Preface
Concepts of Biology is intended for the introductory biology course for non-science majors taught at most two- and four-year colleges. The scope, sequence, and level of the program are designed to match typical course syllabi. This text includes interesting features that make connections between scientific concepts and the everyday world of students. Concepts of Biology conveys the major themes of biology, such as a foundation in evolution, and features a rich and engaging art program.

Welcome to *Concepts of Biology*, an OpenStax resource. This textbook has been created with several goals in mind: accessibility, customization, and student engagement—all while encouraging students toward high levels of academic scholarship. Instructors and students alike will find that this textbook offers a strong introduction to biology in an accessible format.

About OpenStax

OpenStax is a non-profit organization committed to improving student access to quality learning materials. Our free textbooks are developed and peer-reviewed by educators to ensure they are readable, accurate, and meet the scope and sequence requirements of today's college courses. Unlike traditional textbooks, OpenStax resources live online and are owned by the community of educators using them. Through our partnerships with companies and foundations committed to reducing costs for students, OpenStax is working to improve access to higher education for all. OpenStax is an initiative of Rice University and is made possible through the generous support of several philanthropic foundations.

About OpenStax's Resources

OpenStax resources provide quality academic instruction. Three key features set our materials apart from others: they can be customized by instructors for each class, they are a "living" resource that grows online through contributions from science educators, and they are available free or for minimal cost.

Customization

OpenStax learning resources are designed to be customized for each course. Our textbooks provide a solid foundation on which instructors can build, and our resources are conceived and written with flexibility in mind. Instructors can select the sections most relevant to their curricula and create a textbook that speaks directly to the needs of their classes and student body. Teachers are encouraged to expand on existing examples by adding unique context via geographically localized applications and topical connections.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit your book page on openstax.org for more information.

Curation

To broaden access and encourage community curation, *Concepts of Biology* is “open source” licensed under a Creative Commons Attribution (CC-BY) license. The scientific community is invited to submit examples, emerging research, and other feedback to enhance and strengthen the material and keep it current and relevant for today’s students. You can submit your suggestions to Support@OpenStax.org.

Cost

Our textbooks are available for free online, and in low-cost print and e-book editions.

About *Concepts of Biology*

Concepts of Biology is designed for the single-semester introduction to biology course for non-science majors, which for many students is their

only college-level science course. As such, this course represents an important opportunity for students to develop the necessary knowledge, tools, and skills to make informed decisions as they continue with their lives. Rather than being mired down with facts and vocabulary, the typical non-science major student needs information presented in a way that is easy to read and understand. Even more importantly, the content should be meaningful. Students do much better when they understand why biology is relevant to their everyday lives. For these reasons, *Concepts of Biology* is grounded on an evolutionary basis and includes exciting features that highlight careers in the biological sciences and everyday applications of the concepts at hand. We also strive to show the interconnectedness of topics within this extremely broad discipline. In order to meet the needs of today's instructors and students, we maintain the overall organization and coverage found in most syllabi for this course. A strength of *Concepts of Biology* is that instructors can customize the book, adapting it to the approach that works best in their classroom. Concepts of Biology also includes an innovative art program that incorporates critical thinking and clicker questions to help students understand—and apply—key concepts.

Coverage and Scope

Our *Concepts of Biology* textbook adheres to the scope and sequence of most one-semester non-majors courses nationwide. We also strive to make biology, as a discipline, interesting and accessible to students. In addition to a comprehensive coverage of core concepts and foundational research, we have incorporated features that draw learners into the discipline in meaningful ways. Our scope of content was developed after surveying over a hundred biology professors and listening to their coverage needs. We provide a thorough treatment of biology's fundamental concepts with a scope that is manageable for instructors and students alike.

> Unit 1: **The Cellular Foundation of Life**. Our opening unit introduces students to the sciences, including the process of science and the underlying concepts from the physical sciences that provide a framework within which learners comprehend biological processes.

Additionally, students will gain solid understanding of the structures, functions, and processes of the most basic unit of life: the cell.
Unit 2: **Cell Division and Genetics**. Our genetics unit takes learners from the foundations of cellular reproduction to the experiments that revealed the basis of genetics and laws of inheritance.
Unit 3: **Molecular Biology and Biotechnology**. Students will learn the intricacies of DNA, protein synthesis, and gene regulation and current applications of biotechnology and genomics.
Unit 4: **Evolution and the Diversity of Life**. The core concepts of evolution are discussed in this unit with examples illustrating evolutionary processes. Additionally, the evolutionary basis of biology reappears throughout the textbook in general discussion and is reinforced through special call-out features highlighting specific evolution-based topics. The diversity of life is explored with detailed study of various organisms and discussion of emerging phylogenetic relationships between and among bacteria, protist kingdoms, fungi, plants, and animals.
Unit 5: **Animal Structure and Function**. An introduction to the form and function of the animal body is followed by chapters on the immune system and animal development. This unit touches on the biology of all organisms while maintaining an engaging focus on human anatomy and physiology that helps students connect to the topics.
Unit 6: **Ecology**. Ecological concepts are broadly covered in this unit, with features highlighting localized, real-world issues of conservation and biodiversity.

Pedagogical Foundation and Features

Because of the impact science has on students and society, an important goal of science education is to achieve a scientifically literate population that consistently makes informed decisions. Scientific literacy transcends a basic understanding of scientific principles and processes to include the ability to make sense of the myriad instances where people encounter science in day-to-day life. Thus, a scientifically literate person is one who uses science content knowledge to make informed decisions, either

personally or socially, about topics or issues that have a connection with science. Concepts of Biology is grounded on a solid scientific base and designed to promote scientific literacy. Throughout the text, you will find features that engage the students in scientific inquiry by taking selected topics a step further.

- **Evolution in Action** features uphold the importance of evolution to all biological study through discussions like “Global Decline of Coral Reefs” and “The Red Queen Hypothesis.”
- **Career in Action** features present information on a variety of careers in the biological sciences, introducing students to the educational requirements and day-to-day work life of a variety of professions, such as forensic scientists, registered dietitians, and biogeographers.
- **Biology in Action** features tie biological concepts to emerging issues and discuss science in terms of everyday life. Topics include “Invasive Species” and “Photosynthesis at the Grocery Store.”

Art and Animations that Engage

Our art program takes a straightforward approach designed to help students learn the concepts of biology through simple, effective illustrations, photos, and micrographs. Concepts of Biology also incorporates links to relevant animations and interactive exercises that help bring biology to life for students.

- **Art Connection** features call out core figures in each chapter for student attention. Questions about key figures, including clicker questions that can be used in the classroom, engage students’ critical thinking and analytical abilities to ensure their genuine understanding of the concept at hand.
- **Concepts in Action** features direct students to online interactive exercises and animations to add a fuller context and examples to core content.

About Our Team

Concepts of Biology would not be possible if not for the tremendous contributions of the authors and community reviewing team

Senior Contributing Authors

Samantha Fowler, Clayton State University
Rebecca Roush, Sandhills Community College
James Wise, Hampton University

Contributing Authors and Reviewers

Mark Belk, Brigham Young University
Lisa Boggs, Southwestern Oklahoma State University
Sherryl Broverman, Duke University
David Byres, Florida State College at Jacksonville
Aaron Cassill, The University of Texas at San Antonio
Karen Champ, College of Central Florida
Sue Chaplin, University of St. Thomas
Diane Day, Clayton State University
Jean DeSaix, University of North Carolina at Chapel Hill
David Hunnicutt, St. Norbert College
Barbara Kuehner, Hawaii Community College
Brenda Leady, University of Toledo
Bernie Marcus, Genesee Community College
Flora Mhlanga, Lipscomb University
Madeline Mignone, Dominican College
Elizabeth Nash, Long Beach City College
Mark Newton, San Jose City College
Diana Oliveras, University of Colorado Boulder
Ann Paterson, Williams Baptist College
Joel Piperberg, Millersville University
Nick Reeves, Mt. San Jacinto College
Ann Reisenauer, San Jose State University
Lynn Rumfelt, Gordon College

Michael Rutledge, Middle Tennessee State University
Edward Saiff, Ramapo College of New Jersey
Brian Shmaefsky, Kingwood College
Gary Shultz, Marshall University
Donald Slish, SUNY Plattsburgh
Anh-Hue Tu, Georgia Southwestern State University
Elena Zoubina, Bridgewater State University

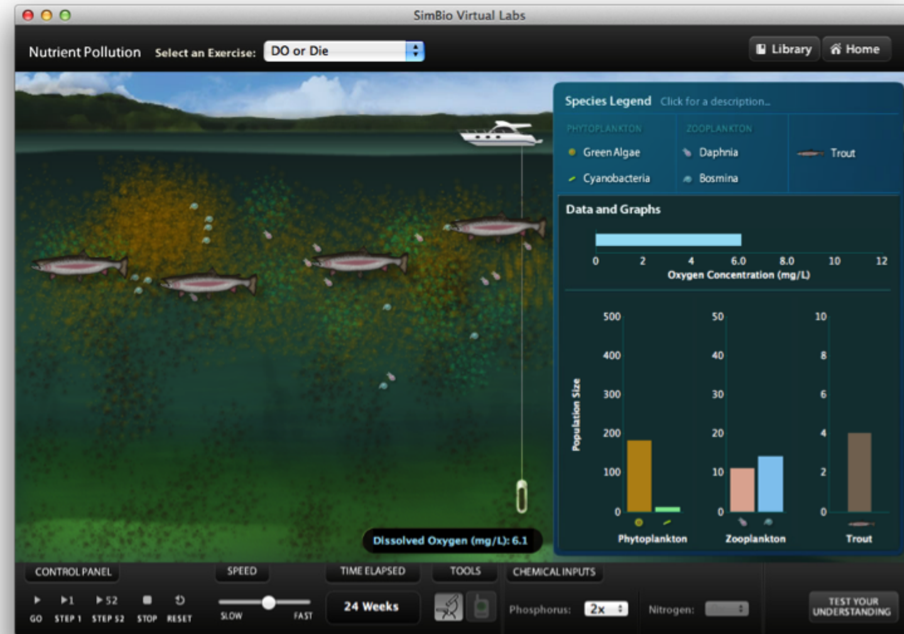
Learning Resources

Wiley Plus for Biology-Fall 2013 Pilot
WileyPLUS provides an engaging online environment for effective teaching and learning. WileyPLUS builds students' confidence because it takes the guesswork out of studying by providing a clear roadmap; what to do, how to do it, and if they did it right. With WileyPLUS, students take more initiative. Therefore, the course has a greater impact on their learning experience. Adaptive tools provide students with a personal, adaptive learning experience so they can build their proficiency on topics and use their study time most effectively. Please let us know if you would like to participate in a Fall 2013 Pilot.

Concepts of Biology Powerpoint Slides (faculty only)
The PowerPoint slides are based on the extensive illustrations from College Physics. They can be edited, incorporated into lecture notes, and you are free to share with anyone in the community. This is a restricted item requiring faculty registration. NOTE: This file is very large and may take some time to download.

SimBio (Laboratory)
SimBio's interactive modules (virtual labs and interactive tutorials and chapters) provide engaging, discovery-based learning tools that complement many of the chapters of Concepts of Biology. SimBio is best known for their EcoBeaker® and EvoBeaker® suites of simulated ecology and evolution laboratories that guide students through the "discovery" of important concepts via a mix of structured and open-ended experimentation on simulated systems. In response to popular demand, SimBio has begun applying the same powerful approaches to topics in cell biology, genetics,

and neurobiology. All of SimBio's modules include instant-feedback questions that enhance student comprehension and auto-graded questions that facilitate implementation.



Introduction
class="introduction"

This NASA image is a composite of several satellite-based views of Earth. To make the whole-Earth image, NASA scientists combine observations of different parts of the planet. (credit: modification of work by NASA)

Viewed from space, Earth ([link]) offers few clues about the diversity of life forms that reside there. The first forms of life on Earth are thought to have been microorganisms that existed for billions of years before plants and animals appeared. The mammals, birds, and flowers so familiar to us are all relatively recent, originating 130 to 200 million years ago. Humans have inhabited this planet for only the last 2.5 million years, and only in the last 200,000 years have humans started looking like we do today.

Themes and Concepts of Biology
By the end of this section, you will be able to:

- Identify and describe the properties of life
- Describe the levels of organization among living things
- List examples of different sub disciplines in biology

Biology is the science that studies life. What exactly is life? This may sound like a silly question with an obvious answer, but it is not easy to define life. For example, a branch of biology called virology studies viruses, which exhibit some of the characteristics of living entities but lack others. It turns out that although viruses can attack living organisms, cause diseases, and even reproduce, they do not meet the criteria that biologists use to define life.

From its earliest beginnings, biology has wrestled with four questions: What are the shared properties that make something “alive”? How do those various living things function? When faced with the remarkable diversity of life, how do we organize the different kinds of organisms so that we can better understand them? And, finally—what biologists ultimately seek to understand—how did this diversity arise and how is it continuing? As new organisms are discovered every day, biologists continue to seek answers to these and other questions.

Properties of Life

All groups of living organisms share several key characteristics or functions: order, sensitivity or response to stimuli, reproduction, adaptation, growth and development, regulation, homeostasis, and energy processing. When viewed together, these eight characteristics serve to define life.

Order

Organisms are highly organized structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex. Inside each cell, atoms make up molecules. These in turn make up cell

components or organelles. Multicellular organisms, which may consist of millions of individual cells, have an advantage over single-celled organisms in that their cells can be specialized to perform specific functions, and even sacrificed in certain situations for the good of the organism as a whole. How these specialized cells come together to form organs such as the heart, lung, or skin in organisms like the toad shown in [link] will be discussed later.

A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems. (credit: "Ivengo(RUS)"/Wikimedia Commons)

Sensitivity or Response to Stimuli

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light or respond to touch ([link]). Even tiny bacteria can move toward or away from chemicals (a process called chemotaxis) or light (phototaxis). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.

The leaves of this sensitive plant (*Mimosa pudica*) will instantly droop and fold when touched. After a few minutes, the plant returns to its normal state. (credit: Alex Lomas)

Note:
Concept in Action

Watch this video to see how the sensitive plant responds to a touch stimulus.

Reproduction

Single-celled organisms reproduce by first duplicating their DNA, which is the genetic material, and then dividing it equally as the cell prepares to divide to form two new cells. Many multicellular organisms (those made up of more than one cell) produce specialized reproductive cells that will form new individuals. When reproduction occurs, DNA containing genes is passed along to an organism’s offspring. These genes are the reason that the offspring will belong to the same species and will have characteristics similar to the parent, such as fur color and blood type.

Adaptation

All living organisms exhibit a “fit” to their environment. Biologists refer to this fit as adaptation and it is a consequence of evolution by natural selection, which operates in every lineage of reproducing organisms. Examples of adaptations are diverse and unique, from heat-resistant Archaea that live in boiling hot springs to the tongue length of a nectar-feeding moth that matches the size of the flower from which it feeds. All adaptations enhance the reproductive potential of the individual exhibiting them, including their ability to survive to reproduce. Adaptations are not constant. As an environment changes, natural selection causes the characteristics of the individuals in a population to track those changes.

Growth and Development

Organisms grow and develop according to specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species’ young ([link]) will grow up to exhibit many of the same characteristics as its parents.

Although no two look alike, these kittens have inherited genes from both parents and share many of the same characteristics. (credit: Pieter & Renée Lanser)

Regulation

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, such as the transport of nutrients, response to stimuli, and coping with environmental stresses. For example, organ systems such as the digestive or circulatory systems perform specific functions like carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

Homeostasis

To function properly, cells require appropriate conditions such as proper temperature, pH, and concentrations of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to

maintain internal conditions within a narrow range almost constantly, despite environmental changes, through a process called **homeostasis** or “steady state”—the ability of an organism to maintain constant internal conditions. For example, many organisms regulate their body temperature in a process known as thermoregulation. Organisms that live in cold climates, such as the polar bear ([link]), have body structures that help them withstand low temperatures and conserve body heat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

Polar bears and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin. (credit: "longhorndave"/Flickr)

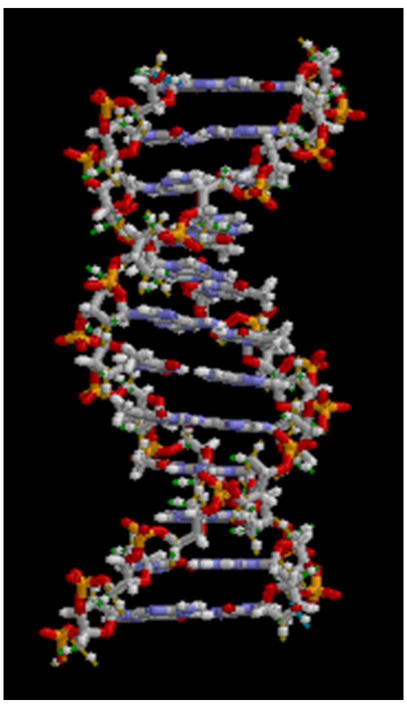
Energy Processing

All organisms (such as the California condor shown in [link]) use a source of energy for their metabolic activities. Some organisms capture energy from the Sun and convert it into chemical energy in food; others use chemical energy from molecules they take in.

A lot of energy is required for a California condor to fly. Chemical energy derived from food is used to power flight. California condors are an endangered species; scientists have strived to place a wing tag on each bird to help them identify and locate each individual bird. (credit: Pacific Southwest Region U.S. Fish and Wildlife)

Levels of Organization of Living Things

Living things are highly organized and structured, following a hierarchy on a scale from small to large. The **atom** is the smallest and most fundamental unit of matter. It consists of a nucleus surrounded by electrons. Atoms form molecules. A **molecule** is a chemical structure consisting of at least two atoms held together by a chemical bond. Many molecules that are biologically important are **macromolecules**, large molecules that are typically formed by combining smaller units called monomers. An example of a macromolecule is deoxyribonucleic acid (DNA) ([link]), which contains the instructions for the functioning of the organism that contains it.



A molecule, like this large DNA molecule, is composed of atoms. (credit: "Brian0918"/Wikimedia Commons)

Note:
Concept in Action

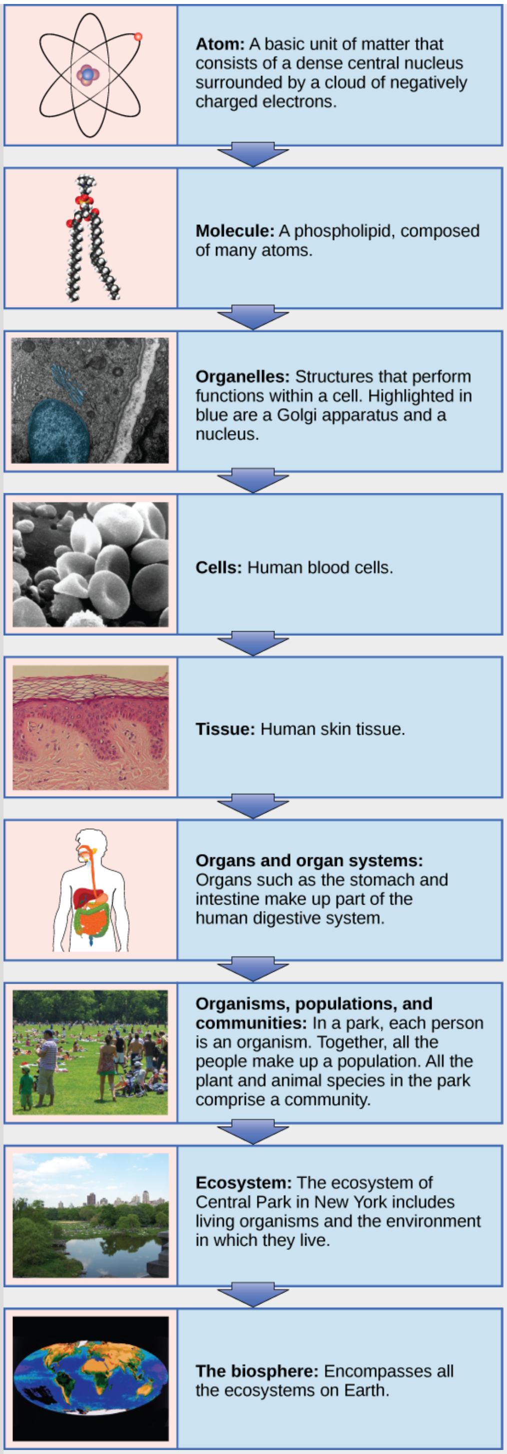
To see an animation of this DNA molecule, click here.

Some cells contain aggregates of macromolecules surrounded by membranes; these are called **organelles**. Organelles are small structures that exist within cells and perform specialized functions. All living things are made of cells; the **cell** itself is the smallest fundamental unit of structure and function in living organisms. (This requirement is why viruses are not considered living: they are not made of cells. To make new viruses, they have to invade and hijack a living cell; only then can they obtain the materials they need to reproduce.) Some organisms consist of a single cell and others are multicellular. Cells are classified as prokaryotic or eukaryotic. **Prokaryotes** are single-celled organisms that lack organelles surrounded by a membrane and do not have nuclei surrounded by nuclear membranes; in contrast, the cells of **eukaryotes** do have membrane-bound organelles and nuclei.

In most multicellular organisms, cells combine to make **tissues**, which are groups of similar cells carrying out the same function. **Organs** are collections of tissues grouped together based on a common function. Organs are present not only in animals but also in plants. An **organ system** is a higher level of organization that consists of functionally related organs. For example vertebrate animals have many organ systems, such as the

circulatory system that transports blood throughout the body and to and from the lungs; it includes organs such as the heart and blood vessels. **Organisms** are individual living entities. For example, each tree in a forest is an organism. Single-celled prokaryotes and single-celled eukaryotes are also considered organisms and are typically referred to as microorganisms.

Note:
Art Connection



From an atom to the entire Earth, biology examines all aspects of life. (credit "molecule": modification of work by

Jane Whitney; credit "organelles": modification of work by Louisa Howard; credit "cells": modification of work by Bruce Wetzel, Harry Schaefer, National Cancer Institute; credit "tissue": modification of work by "Kilbad"/Wikimedia Commons; credit "organs": modification of work by Mariana Ruiz Villareal, Joaquim Alves Gaspar; credit "organisms": modification of work by Peter Dutton; credit "ecosystem": modification of work by "gigi4791"/Flickr; credit "biosphere": modification of work by NASA)

Which of the following statements is false?

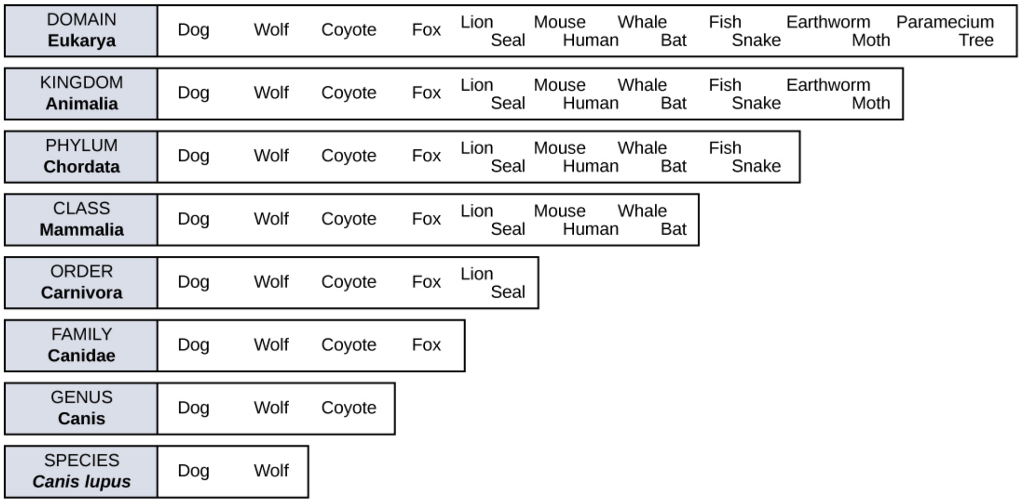
a. Tissues exist within organs which exist within organ systems.
b. Communities exist within populations which exist within ecosystems.
c. Organelles exist within cells which exist within tissues.
d. Communities exist within ecosystems which exist in the biosphere.

All the individuals of a species living within a specific area are collectively called a **population**. For example, a forest may include many white pine trees. All of these pine trees represent the population of white pine trees in this forest. Different populations may live in the same specific area. For example, the forest with the pine trees includes populations of flowering plants and also insects and microbial populations. A **community** is the set of populations inhabiting a particular area. For instance, all of the trees, flowers, insects, and other populations in a forest form the forest's community. The forest itself is an ecosystem. An **ecosystem** consists of all the living things in a particular area together with the abiotic, or non-living, parts of that environment such as nitrogen in the soil or rainwater. At the highest level of organization ([link]), the **biosphere** is the collection of all ecosystems, and it represents the zones of life on Earth. It includes land, water, and portions of the atmosphere.

The Diversity of Life

The science of biology is very broad in scope because there is a tremendous diversity of life on Earth. The source of this diversity is **evolution**, the process of gradual change during which new species arise from older species. Evolutionary biologists study the evolution of living things in everything from the microscopic world to ecosystems.

In the 18th century, a scientist named Carl Linnaeus first proposed organizing the known species of organisms into a hierarchical taxonomy. In this system, species that are most similar to each other are put together within a grouping known as a genus. Furthermore, similar genera (the plural of genus) are put together within a family. This grouping continues until all organisms are collected together into groups at the highest level. The current taxonomic system now has eight levels in its hierarchy, from lowest to highest, they are: species, genus, family, order, class, phylum, kingdom, domain. Thus species are grouped within genera, genera are grouped within families, families are grouped within orders, and so on ([link]).

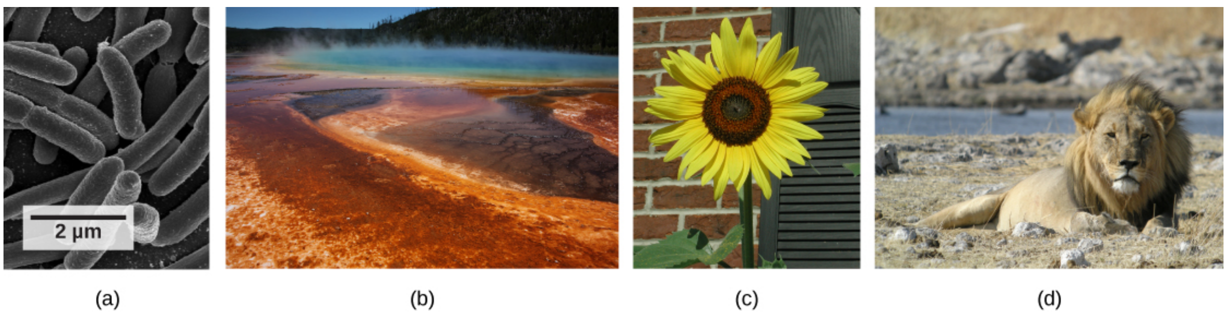


This diagram shows the levels of taxonomic hierarchy for a dog, from the broadest category—domain—to the most specific—species.

The highest level, domain, is a relatively new addition to the system since the 1990s. Scientists now recognize three domains of life, the Eukarya, the Archaea, and the Bacteria. The domain Eukarya contains organisms that have cells with nuclei. It includes the kingdoms of fungi, plants, animals, and several kingdoms of protists. The Archaea, are single-celled organisms without nuclei and include many extremophiles that live in harsh environments like hot springs. The Bacteria are another quite different group of single-celled organisms without nuclei ([link]). Both the Archaea and the Bacteria are prokaryotes, an informal name for cells without nuclei. The recognition in the 1990s that certain “bacteria,” now known as the Archaea, were as different genetically and biochemically from other bacterial cells as they were from eukaryotes, motivated the recommendation to divide life into three domains. This dramatic change in our knowledge of the tree of life demonstrates that classifications are not permanent and will change when new information becomes available.

In addition to the hierarchical taxonomic system, Linnaeus was the first to name organisms using two unique names, now called the binomial naming system. Before Linnaeus, the use of common names to refer to organisms

caused confusion because there were regional differences in these common names. Binomial names consist of the genus name (which is capitalized) and the species name (all lower-case). Both names are set in italics when they are printed. Every species is given a unique binomial which is recognized the world over, so that a scientist in any location can know which organism is being referred to. For example, the North American blue jay is known uniquely as *Cyanocitta cristata*. Our own species is *Homo sapiens*.

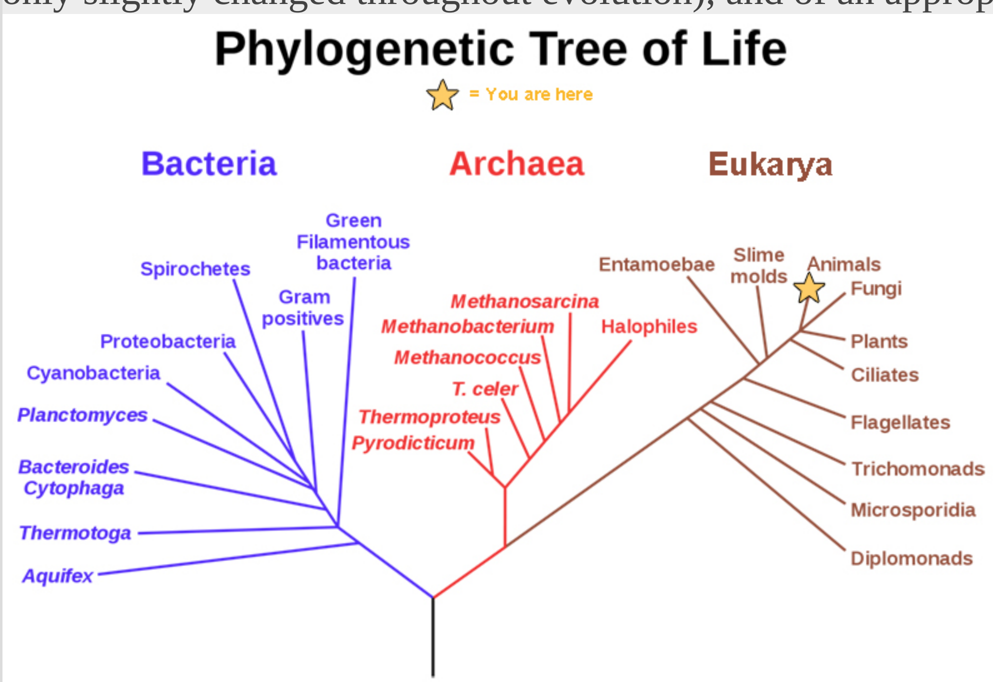


These images represent different domains. The scanning electron micrograph shows (a) bacterial cells belong to the domain Bacteria, while the (b) extremophiles, seen all together as colored mats in this hot spring, belong to domain Archaea. Both the (c) sunflower and (d) lion are part of domain Eukarya. (credit a: modification of work by Rocky Mountain Laboratories, NIAID, NIH; credit b: modification of work by Steve Jurvetson; credit c: modification of work by Michael Arrighi; credit d: modification of work by Frank Vassen)

Note:
Evolution in Action
Carl Woese and the Phylogenetic Tree
The evolutionary relationships of various life forms on Earth can be summarized in a phylogenetic tree. A **phylogenetic tree** is a diagram showing the evolutionary relationships among biological species based on

similarities and differences in genetic or physical traits or both. A phylogenetic tree is composed of branch points, or nodes, and branches. The internal nodes represent ancestors and are points in evolution when, based on scientific evidence, an ancestor is thought to have diverged to form two new species. The length of each branch can be considered as estimates of relative time.

In the past, biologists grouped living organisms into five kingdoms: animals, plants, fungi, protists, and bacteria. The pioneering work of American microbiologist Carl Woese in the early 1970s has shown, however, that life on Earth has evolved along three lineages, now called domains—Bacteria, Archaea, and Eukarya. Woese proposed the domain as a new taxonomic level and Archaea as a new domain, to reflect the new phylogenetic tree ([link]). Many organisms belonging to the Archaea domain live under extreme conditions and are called extremophiles. To construct his tree, Woese used genetic relationships rather than similarities based on morphology (shape). Various genes were used in phylogenetic studies. Woese's tree was constructed from comparative sequencing of the genes that are universally distributed, found in some slightly altered form in every organism, conserved (meaning that these genes have remained only slightly changed throughout evolution), and of an appropriate length.



This phylogenetic tree was constructed by microbiologist

Carl Woese using genetic relationships. The tree shows the separation of living organisms into three domains: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are organisms without a nucleus or other organelles surrounded by a membrane and, therefore, are prokaryotes. (credit: modification of work by Eric Gaba)

Branches of Biological Study

The scope of biology is broad and therefore contains many branches and sub disciplines. Biologists may pursue one of those sub disciplines and work in a more focused field. For instance, molecular biology studies biological processes at the molecular level, including interactions among molecules such as DNA, RNA, and proteins, as well as the way they are regulated. Microbiology is the study of the structure and function of microorganisms. It is quite a broad branch itself, and depending on the subject of study, there are also microbial physiologists, ecologists, and geneticists, among others.

Another field of biological study, neurobiology, studies the biology of the nervous system, and although it is considered a branch of biology, it is also recognized as an interdisciplinary field of study known as neuroscience. Because of its interdisciplinary nature, this sub discipline studies different functions of the nervous system using molecular, cellular, developmental, medical, and computational approaches.

Researchers work on excavating dinosaur fossils at a site in Castellón, Spain. (credit: Mario Modesto)

Paleontology, another branch of biology, uses fossils to study life's history ([link]). Zoology and botany are the study of animals and plants, respectively. Biologists can also specialize as biotechnologists, ecologists, or physiologists, to name just a few areas. Biotechnologists apply the knowledge of biology to create useful products. Ecologists study the interactions of organisms in their environments. Physiologists study the workings of cells, tissues and organs. This is just a small sample of the many fields that biologists can pursue. From our own bodies to the world we live in, discoveries in biology can affect us in very direct and important ways. We depend on these discoveries for our health, our food sources, and the benefits provided by our ecosystem. Because of this, knowledge of biology can benefit us in making decisions in our day-to-day lives.

The development of technology in the twentieth century that continues today, particularly the technology to describe and manipulate the genetic material, DNA, has transformed biology. This transformation will allow biologists to continue to understand the history of life in greater detail, how the human body works, our human origins, and how humans can survive as

a species on this planet despite the stresses caused by our increasing numbers. Biologists continue to decipher huge mysteries about life suggesting that we have only begun to understand life on the planet, its history, and our relationship to it. For this and other reasons, the knowledge of biology gained through this textbook and other printed and electronic media should be a benefit in whichever field you enter.

Note:
Careers in Action
Forensic Scientist
Forensic science is the application of science to answer questions related to the law. Biologists as well as chemists and biochemists can be forensic scientists. Forensic scientists provide scientific evidence for use in courts, and their job involves examining trace material associated with crimes. Interest in forensic science has increased in the last few years, possibly because of popular television shows that feature forensic scientists on the job. Also, the development of molecular techniques and the establishment of DNA databases have updated the types of work that forensic scientists can do. Their job activities are primarily related to crimes against people such as murder, rape, and assault. Their work involves analyzing samples such as hair, blood, and other body fluids and also processing DNA ([link]) found in many different environments and materials. Forensic scientists also analyze other biological evidence left at crime scenes, such as insect parts or pollen grains. Students who want to pursue careers in forensic science will most likely be required to take chemistry and biology courses as well as some intensive math courses.

This forensic scientist works in a DNA extraction room at the U.S. Army Criminal Investigation Laboratory. (credit: U.S. Army CID Command Public Affairs)

Section Summary

Biology is the science of life. All living organisms share several key properties such as order, sensitivity or response to stimuli, reproduction, adaptation, growth and development, regulation, homeostasis, and energy processing. Living things are highly organized following a hierarchy that includes atoms, molecules, organelles, cells, tissues, organs, and organ systems. Organisms, in turn, are grouped as populations, communities, ecosystems, and the biosphere. Evolution is the source of the tremendous biological diversity on Earth today. A diagram called a phylogenetic tree can be used to show evolutionary relationships among organisms. Biology is very broad and includes many branches and sub disciplines. Examples

include molecular biology, microbiology, neurobiology, zoology, and botany, among others.

Art Connections

Exercise:

Problem: [link] Which of the following statements is false?

A. Tissues exist within organs which exist within organ systems.
B. Communities exist within populations which exist within ecosystems.
C. Organelles exist within cells which exist within tissues.
D. Communities exist within ecosystems which exist in the biosphere.

Solution:

[link] B

Multiple Choice

Exercise:

Problem:

The smallest unit of biological structure that meets the functional requirements of “living” is the _________.

a. organ
b. organelle
c. cell
d. macromolecule

Solution:

C

Exercise:

Problem:

Which of the following sequences represents the hierarchy of biological organization from the most complex to the least complex level?

a. organelle, tissue, biosphere, ecosystem, population
b. organ, organism, tissue, organelle, molecule
c. organism, community, biosphere, molecule, tissue, organ
d. biosphere, ecosystem, community, population, organism

Solution:

D

Free Response

Exercise:

Problem:

Using examples, explain how biology can be studied from a microscopic approach to a global approach.

Solution:

Researchers can approach biology from the smallest to the largest, and everything in between. For instance, an ecologist may study a population of individuals, the population’s community, the community’s ecosystem, and the ecosystem’s part in the biosphere. When studying an individual organism, a biologist could examine the cell and its organelles, the tissues that the cells make up, the organs and their respective organ systems, and the sum total—the organism itself.

Glossary

atom
: a basic unit of matter that cannot be broken down by normal chemical reactions

biology
: the study of living organisms and their interactions with one another and their environments

biosphere
: a collection of all ecosystems on Earth

cell
: the smallest fundamental unit of structure and function in living things

community
: a set of populations inhabiting a particular area

ecosystem
: all living things in a particular area together with the abiotic, nonliving parts of that environment

eukaryote
: an organism with cells that have nuclei and membrane-bound organelles

evolution
: the process of gradual change in a population that can also lead to new species arising from older species

homeostasis
: the ability of an organism to maintain constant internal conditions

macromolecule
: a large molecule typically formed by the joining of smaller molecules

molecule
: a chemical structure consisting of at least two atoms held together by a chemical bond

organ
: a structure formed of tissues operating together to perform a common function

organ system
: the higher level of organization that consists of functionally related organs

organelle
: a membrane-bound compartment or sac within a cell

organism
: an individual living entity

phylogenetic tree
: a diagram showing the evolutionary relationships among biological species based on similarities and differences in genetic or physical traits or both

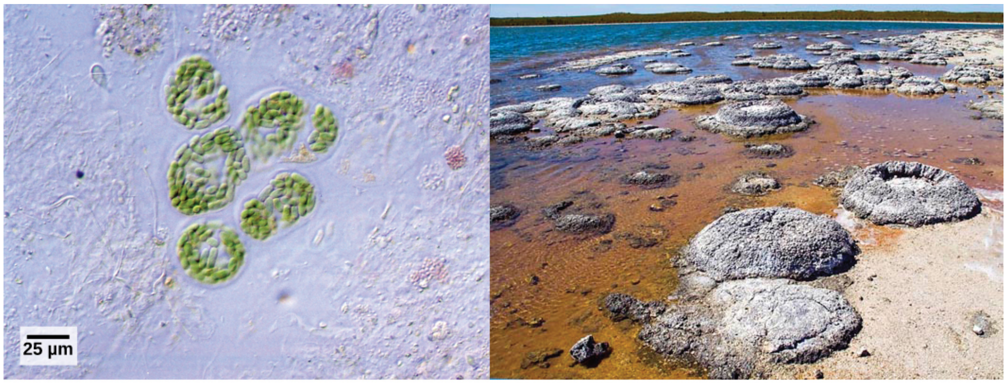
population
: all individuals within a species living within a specific area

prokaryote
: a unicellular organism that lacks a nucleus or any other membrane-bound organelle

tissue
: a group of similar cells carrying out the same function

The Process of Science
By the end of this section, you will be able to:

- Identify the shared characteristics of the natural sciences
- Understand the process of scientific inquiry
- Compare inductive reasoning with deductive reasoning
- Describe the goals of basic science and applied science

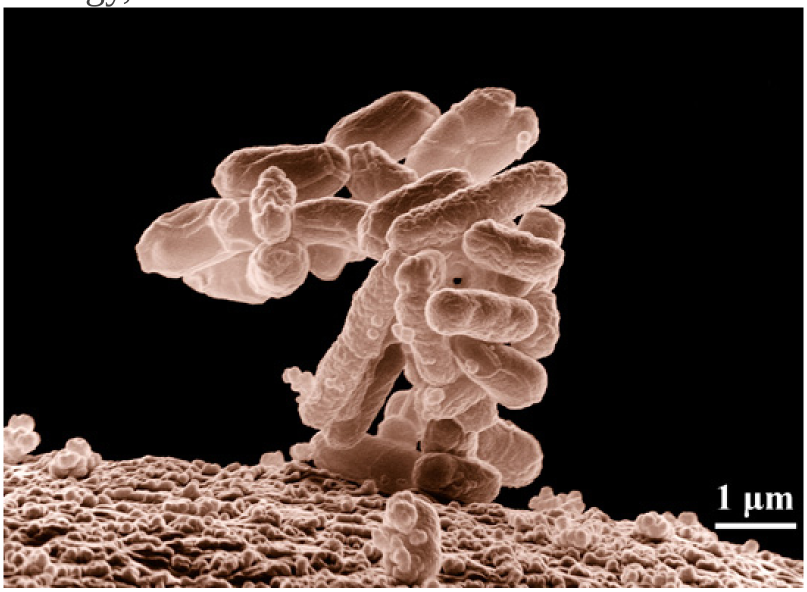


(a) (b)

Formerly called blue-green algae, the (a) cyanobacteria seen through a light microscope are some of Earth's oldest life forms. These (b) stromatolites along the shores of Lake Thetis in Western Australia are ancient structures formed by the layering of cyanobacteria in shallow waters. (credit a: modification of work by NASA; scale-bar data from Matt Russell; credit b: modification of work by Ruth Ellison)

Like geology, physics, and chemistry, biology is a science that gathers knowledge about the natural world. Specifically, biology is the study of life. The discoveries of biology are made by a community of researchers who work individually and together using agreed-on methods. In this sense, biology, like all sciences is a social enterprise like politics or the arts. The methods of science include careful observation, record keeping, logical and mathematical reasoning, experimentation, and submitting conclusions to the scrutiny of others. Science also requires considerable imagination and

creativity; a well-designed experiment is commonly described as elegant, or beautiful. Like politics, science has considerable practical implications and some science is dedicated to practical applications, such as the prevention of disease (see [link]). Other science proceeds largely motivated by curiosity. Whatever its goal, there is no doubt that science, including biology, has transformed human existence and will continue to do so.



Biologists may choose to study *Escherichia coli* (*E. coli*), a bacterium that is a normal resident of our digestive tracts but which is also sometimes responsible for disease outbreaks. In this micrograph, the bacterium is visualized using a scanning electron microscope and digital colorization. (credit: Eric Erbe; digital colorization by Christopher Pooley, USDA-ARS)

The Nature of Science

Biology is a science, but what exactly is science? What does the study of biology share with other scientific disciplines? **Science** (from the Latin

scientia, meaning "knowledge") can be defined as knowledge about the natural world.

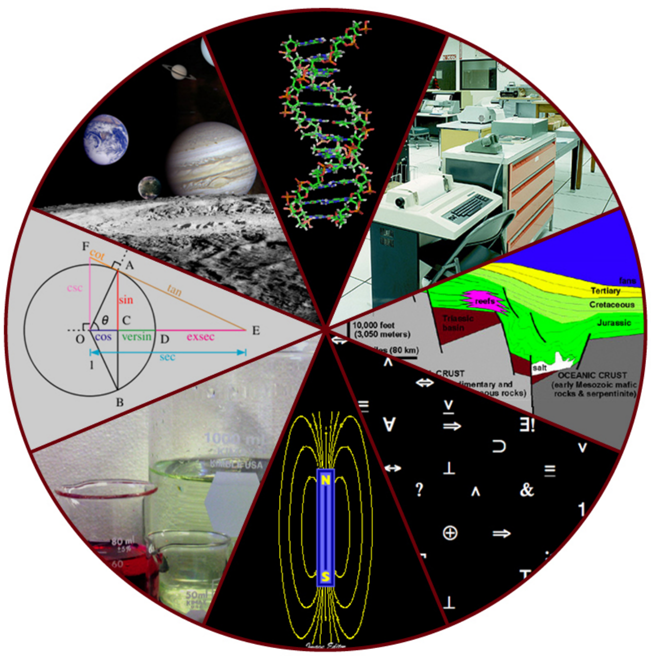
Science is a very specific way of learning, or knowing, about the world. The history of the past 500 years demonstrates that science is a very powerful way of knowing about the world; it is largely responsible for the technological revolutions that have taken place during this time. There are however, areas of knowledge and human experience that the methods of science cannot be applied to. These include such things as answering purely moral questions, aesthetic questions, or what can be generally categorized as spiritual questions. Science cannot investigate these areas because they are outside the realm of material phenomena, the phenomena of matter and energy, and cannot be observed and measured.

The **scientific method** is a method of research with defined steps that include experiments and careful observation. The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses. A **hypothesis** is a suggested explanation for an event, which can be tested. Hypotheses, or tentative explanations, are generally produced within the context of a **scientific theory**. A scientific theory is a generally accepted, thoroughly tested and confirmed explanation for a set of observations or phenomena. Scientific theory is the foundation of scientific knowledge. In addition, in many scientific disciplines (less so in biology) there are **scientific laws**, often expressed in mathematical formulas, which describe how elements of nature will behave under certain specific conditions. There is not an evolution of hypotheses through theories to laws as if they represented some increase in certainty about the world. Hypotheses are the day-to-day material that scientists work with and they are developed within the context of theories. Laws are concise descriptions of parts of the world that are amenable to formulaic or mathematical description.

Natural Sciences

What would you expect to see in a museum of natural sciences? Frogs? Plants? Dinosaur skeletons? Exhibits about how the brain functions? A

planetarium? Gems and minerals? Or maybe all of the above? Science includes such diverse fields as astronomy, biology, computer sciences, geology, logic, physics, chemistry, and mathematics ([link]). However, those fields of science related to the physical world and its phenomena and processes are considered **natural sciences**. Thus, a museum of natural sciences might contain any of the items listed above.



Some fields of science include astronomy, biology, computer science, geology, logic, physics, chemistry, and mathematics. (credit: "Image Editor"/Flickr)

There is no complete agreement when it comes to defining what the natural sciences include. For some experts, the natural sciences are astronomy, biology, chemistry, earth science, and physics. Other scholars choose to divide natural sciences into **life sciences**, which study living things and include biology, and **physical sciences**, which study nonliving matter and

include astronomy, physics, and chemistry. Some disciplines such as biophysics and biochemistry build on two sciences and are interdisciplinary.

Scientific Inquiry

One thing is common to all forms of science: an ultimate goal "to know." Curiosity and inquiry are the driving forces for the development of science. Scientists seek to understand the world and the way it operates. Two methods of logical thinking are used: inductive reasoning and deductive reasoning.

Inductive reasoning is a form of logical thinking that uses related observations to arrive at a general conclusion. This type of reasoning is common in descriptive science. A life scientist such as a biologist makes observations and records them. These data can be qualitative (descriptive) or quantitative (consisting of numbers), and the raw data can be supplemented with drawings, pictures, photos, or videos. From many observations, the scientist can infer conclusions (inductions) based on evidence. Inductive reasoning involves formulating generalizations inferred from careful observation and the analysis of a large amount of data. Brain studies often work this way. Many brains are observed while people are doing a task. The part of the brain that lights up, indicating activity, is then demonstrated to be the part controlling the response to that task.

Deductive reasoning or deduction is the type of logic used in hypothesis-based science. In deductive reasoning, the pattern of thinking moves in the opposite direction as compared to inductive reasoning. **Deductive reasoning** is a form of logical thinking that uses a general principle or law to forecast specific results. From those general principles, a scientist can extrapolate and predict the specific results that would be valid as long as the general principles are valid. For example, a prediction would be that if the climate is becoming warmer in a region, the distribution of plants and animals should change. Comparisons have been made between distributions in the past and the present, and the many changes that have been found are consistent with a warming climate. Finding the change in distribution is evidence that the climate change conclusion is a valid one.

Both types of logical thinking are related to the two main pathways of scientific study: descriptive science and hypothesis-based science. **Descriptive** (or discovery) **science** aims to observe, explore, and discover, while **hypothesis-based science** begins with a specific question or problem and a potential answer or solution that can be tested. The boundary between these two forms of study is often blurred, because most scientific endeavors combine both approaches. Observations lead to questions, questions lead to forming a hypothesis as a possible answer to those questions, and then the hypothesis is tested. Thus, descriptive science and hypothesis-based science are in continuous dialogue.

Hypothesis Testing

Biologists study the living world by posing questions about it and seeking science-based responses. This approach is common to other sciences as well and is often referred to as the scientific method. The scientific method was used even in ancient times, but it was first documented by England's Sir Francis Bacon (1561–1626) ([link]), who set up inductive methods for scientific inquiry. The scientific method is not exclusively used by biologists but can be applied to almost anything as a logical problem-solving method.

Sir Francis Bacon is credited with being the first to document the scientific method.

The scientific process typically starts with an observation (often a problem to be solved) that leads to a question. Let's think about a simple problem that starts with an observation and apply the scientific method to solve the problem. One Monday morning, a student arrives at class and quickly discovers that the classroom is too warm. That is an observation that also describes a problem: the classroom is too warm. The student then asks a question: "Why is the classroom so warm?"

Recall that a hypothesis is a suggested explanation that can be tested. To solve a problem, several hypotheses may be proposed. For example, one hypothesis might be, "The classroom is warm because no one turned on the air conditioning." But there could be other responses to the question, and therefore other hypotheses may be proposed. A second hypothesis might be,

“The classroom is warm because there is a power failure, and so the air conditioning doesn’t work.”

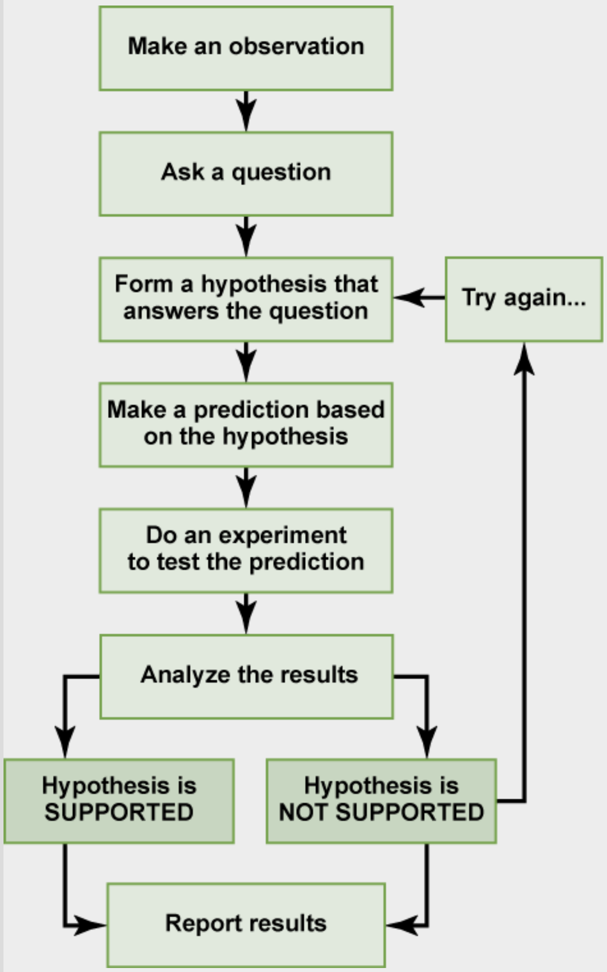
Once a hypothesis has been selected, a prediction may be made. A prediction is similar to a hypothesis but it typically has the format “If . . . then . . . .” For example, the prediction for the first hypothesis might be, “*If* the student turns on the air conditioning, *then* the classroom will no longer be too warm.”

A hypothesis must be testable to ensure that it is valid. For example, a hypothesis that depends on what a bear thinks is not testable, because it can never be known what a bear thinks. It should also be **falsifiable**, meaning that it can be disproven by experimental results. An example of an unfalsifiable hypothesis is “Botticelli’s *Birth of Venus* is beautiful.” There is no experiment that might show this statement to be false. To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. This is important. A hypothesis can be disproven, or eliminated, but it can never be proven. Science does not deal in proofs like mathematics. If an experiment fails to disprove a hypothesis, then we find support for that explanation, but this is not to say that down the road a better explanation will not be found, or a more carefully designed experiment will be found to falsify the hypothesis.

Each experiment will have one or more variables and one or more controls. A **variable** is any part of the experiment that can vary or change during the experiment. A **control** is a part of the experiment that does not change. Look for the variables and controls in the example that follows. As a simple example, an experiment might be conducted to test the hypothesis that phosphate limits the growth of algae in freshwater ponds. A series of artificial ponds are filled with water and half of them are treated by adding phosphate each week, while the other half are treated by adding a salt that is known not to be used by algae. The variable here is the phosphate (or lack of phosphate), the experimental or treatment cases are the ponds with added phosphate and the control ponds are those with something inert added, such as the salt. Just adding something is also a control against the possibility that adding extra matter to the pond has an effect. If the treated ponds show lesser growth of algae, then we have found support for our hypothesis. If

they do not, then we reject our hypothesis. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid ([link]). Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected.

Note:
Art Connection



The scientific method is a series of defined steps that include experiments and careful observation. If a hypothesis is

not supported by data, a new hypothesis can be proposed.

In the example below, the scientific method is used to solve an everyday problem. Which part in the example below is the hypothesis? Which is the prediction? Based on the results of the experiment, is the hypothesis supported? If it is not supported, propose some alternative hypotheses.

1. My toaster doesn’t toast my bread.
2. Why doesn’t my toaster work?
3. There is something wrong with the electrical outlet.
4. If something is wrong with the outlet, my coffeemaker also won’t work when plugged into it.
5. I plug my coffeemaker into the outlet.
6. My coffeemaker works.

In practice, the scientific method is not as rigid and structured as it might at first appear. Sometimes an experiment leads to conclusions that favor a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds. Scientific reasoning is more complex than the scientific method alone suggests.

Basic and Applied Science

The scientific community has been debating for the last few decades about the value of different types of science. Is it valuable to pursue science for the sake of simply gaining knowledge, or does scientific knowledge only have worth if we can apply it to solving a specific problem or bettering our lives? This question focuses on the differences between two types of science: basic science and applied science.

Basic science or “pure” science seeks to expand knowledge regardless of the short-term application of that knowledge. It is not focused on

developing a product or a service of immediate public or commercial value. The immediate goal of basic science is knowledge for knowledge's sake, though this does not mean that in the end it may not result in an application.

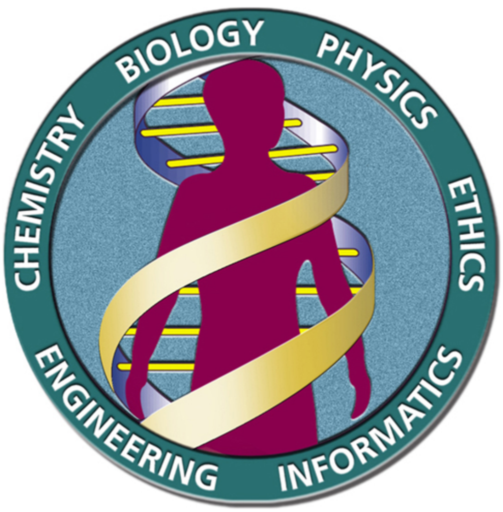
In contrast, **applied science** or "technology," aims to use science to solve real-world problems, making it possible, for example, to improve a crop yield, find a cure for a particular disease, or save animals threatened by a natural disaster. In applied science, the problem is usually defined for the researcher.

Some individuals may perceive applied science as "useful" and basic science as "useless." A question these people might pose to a scientist advocating knowledge acquisition would be, "What for?" A careful look at the history of science, however, reveals that basic knowledge has resulted in many remarkable applications of great value. Many scientists think that a basic understanding of science is necessary before an application is developed; therefore, applied science relies on the results generated through basic science. Other scientists think that it is time to move on from basic science and instead to find solutions to actual problems. Both approaches are valid. It is true that there are problems that demand immediate attention; however, few solutions would be found without the help of the knowledge generated through basic science.

One example of how basic and applied science can work together to solve practical problems occurred after the discovery of DNA structure led to an understanding of the molecular mechanisms governing DNA replication. Strands of DNA, unique in every human, are found in our cells, where they provide the instructions necessary for life. During DNA replication, new copies of DNA are made, shortly before a cell divides to form new cells. Understanding the mechanisms of DNA replication enabled scientists to develop laboratory techniques that are now used to identify genetic diseases, pinpoint individuals who were at a crime scene, and determine paternity. Without basic science, it is unlikely that applied science would exist.

Another example of the link between basic and applied research is the Human Genome Project, a study in which each human chromosome was analyzed and mapped to determine the precise sequence of DNA subunits

and the exact location of each gene. (The gene is the basic unit of heredity; an individual's complete collection of genes is his or her genome.) Other organisms have also been studied as part of this project to gain a better understanding of human chromosomes. The Human Genome Project ([link]) relied on basic research carried out with non-human organisms and, later, with the human genome. An important end goal eventually became using the data for applied research seeking cures for genetically related diseases.



The Human Genome Project was a 13-year collaborative effort among researchers working in several different fields of science. The project was completed in 2003. (credit: the U.S. Department of Energy Genome Programs)

While research efforts in both basic science and applied science are usually carefully planned, it is important to note that some discoveries are made by serendipity, that is, by means of a fortunate accident or a lucky surprise. Penicillin was discovered when biologist Alexander Fleming accidentally left a petri dish of *Staphylococcus* bacteria open. An unwanted mold grew, killing the bacteria. The mold turned out to be *Penicillium*, and a new antibiotic was discovered. Even in the highly organized world of science, luck—when combined with an observant, curious mind—can lead to unexpected breakthroughs.

Reporting Scientific Work

Whether scientific research is basic science or applied science, scientists must share their findings for other researchers to expand and build upon their discoveries. Communication and collaboration within and between sub disciplines of science are key to the advancement of knowledge in science. For this reason, an important aspect of a scientist's work is disseminating results and communicating with peers. Scientists can share results by presenting them at a scientific meeting or conference, but this approach can reach only the limited few who are present. Instead, most scientists present their results in peer-reviewed articles that are published in scientific journals. **Peer-reviewed articles** are scientific papers that are reviewed, usually anonymously by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the scientist's work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.

There are many journals and the popular press that do not use a peer-review system. A large number of online open-access journals, journals with articles available without cost, are now available many of which use rigorous peer-review systems, but some of which do not. Results of any

studies published in these forums without peer review are not reliable and should not form the basis for other scientific work. In one exception, journals may allow a researcher to cite a personal communication from another researcher about unpublished results with the cited author's permission.

Section Summary

Biology is the science that studies living organisms and their interactions with one another and their environments. Science attempts to describe and understand the nature of the universe in whole or in part. Science has many fields; those fields related to the physical world and its phenomena are considered natural sciences.

A hypothesis is a tentative explanation for an observation. A scientific theory is a well-tested and consistently verified explanation for a set of observations or phenomena. A scientific law is a description, often in the form of a mathematical formula, of the behavior of an aspect of nature under certain circumstances. Two types of logical reasoning are used in science. Inductive reasoning uses results to produce general scientific principles. Deductive reasoning is a form of logical thinking that predicts results by applying general principles. The common thread throughout scientific research is the use of the scientific method. Scientists present their results in peer-reviewed scientific papers published in scientific journals.

Science can be basic or applied. The main goal of basic science is to expand knowledge without any expectation of short-term practical application of that knowledge. The primary goal of applied research, however, is to solve practical problems.

Art Connections

Exercise:

Problem:

[link] In the example below, the scientific method is used to solve an everyday problem. Which part in the example below is the hypothesis? Which is the prediction? Based on the results of the experiment, is the hypothesis supported? If it is not supported, propose some alternative hypotheses.

1. My toaster doesn’t toast my bread.
2. Why doesn’t my toaster work?
3. There is something wrong with the electrical outlet.
4. If something is wrong with the outlet, my coffeemaker also won’t work when plugged into it.
5. I plug my coffeemaker into the outlet.
6. My coffeemaker works.

Solution:

[link] The hypothesis is #3 (there is something wrong with the electrical outlet), and the prediction is #4 (if something is wrong with the outlet, then the coffeemaker also won’t work when plugged into the outlet). The original hypothesis is not supported, as the coffee maker works when plugged into the outlet. Alternative hypotheses may include (1) the toaster might be broken or (2) the toaster wasn’t turned on.

Multiple Choice

Exercise:

Problem:

A suggested and testable explanation for an event is called a ________.

a. hypothesis

b. variable
c. theory
d. control

Solution:

A

Exercise:

Problem:

The type of logical thinking that uses related observations to arrive at a general conclusion is called _________.

a. deductive reasoning
b. the scientific method
c. hypothesis-based science
d. inductive reasoning

Solution:

D

Free Response

Exercise:

Problem:

Give an example of how applied science has had a direct effect on your daily life.

Solution:

Answers will vary. One example of how applied science has had a direct effect on daily life is the presence of vaccines. Vaccines to

prevent diseases such polio, measles, tetanus, and even the influenza affect daily life by contributing to individual and societal health.

Glossary

applied science
: a form of science that solves real-world problems

basic science
: science that seeks to expand knowledge regardless of the short-term application of that knowledge

control
: a part of an experiment that does not change during the experiment

deductive reasoning
: a form of logical thinking that uses a general statement to forecast specific results

descriptive science
: a form of science that aims to observe, explore, and find things out

falsifiable
: able to be disproven by experimental results

hypothesis
: a suggested explanation for an event, which can be tested

hypothesis-based science
: a form of science that begins with a specific explanation that is then tested

inductive reasoning
: a form of logical thinking that uses related observations to arrive at a general conclusion

life science

a field of science, such as biology, that studies living things

natural science
: a field of science that studies the physical world, its phenomena, and processes

peer-reviewed article
: a scientific report that is reviewed by a scientist’s colleagues before publication

physical science
: a field of science, such as astronomy, physics, and chemistry, that studies nonliving matter

science
: knowledge that covers general truths or the operation of general laws, especially when acquired and tested by the scientific method

scientific law
: a description, often in the form of a mathematical formula, for the behavior of some aspect of nature under certain specific conditions

scientific method
: a method of research with defined steps that include experiments and careful observation

scientific theory
: a thoroughly tested and confirmed explanation for observations or phenomena

variable
: a part of an experiment that can vary or change

Introduction
class="introduction"

Foods such as bread, fruit, and cheese are rich sources of biological macromolecules. (credit: modification of work by Bengt Nyman)

The elements carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus are the key building blocks of the chemicals found in living things. They form the carbohydrates, nucleic acids, proteins, and lipids (all of which will

be defined later in this chapter) that are the fundamental molecular components of all organisms. In this chapter, we will discuss these important building blocks and learn how the unique properties of the atoms of different elements affect their interactions with other atoms to form the molecules of life.

Food provides an organism with nutrients—the matter it needs to survive. Many of these critical nutrients come in the form of biological macromolecules, or large molecules necessary for life. These macromolecules are built from different combinations of smaller organic molecules. What specific types of biological macromolecules do living things require? How are these molecules formed? What functions do they serve? In this chapter, we will explore these questions.

The Building Blocks of Molecules

By the end of this section, you will be able to:

- Describe matter and elements
- Describe the interrelationship between protons, neutrons, and electrons, and the ways in which electrons can be donated or shared between atoms

At its most fundamental level, life is made up of matter. **Matter** occupies space and has mass. All matter is composed of **elements**, substances that cannot be broken down or transformed chemically into other substances. Each element is made of atoms, each with a constant number of protons and unique properties. A total of 118 elements have been defined; however, only 92 occur naturally, and fewer than 30 are found in living cells. The remaining 26 elements are unstable and, therefore, do not exist for very long or are theoretical and have yet to be detected.

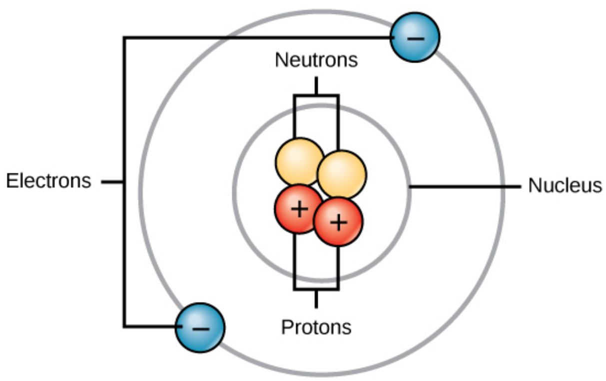
Each element is designated by its chemical symbol (such as H, N, O, C, and Na), and possesses unique properties. These unique properties allow elements to combine and to bond with each other in specific ways.

Atoms

An atom is the smallest component of an element that retains all of the chemical properties of that element. For example, one hydrogen atom has all of the properties of the element hydrogen, such as it exists as a gas at room temperature, and it bonds with oxygen to create a water molecule. Hydrogen atoms cannot be broken down into anything smaller while still retaining the properties of hydrogen. If a hydrogen atom were broken down into subatomic particles, it would no longer have the properties of hydrogen.

At the most basic level, all organisms are made of a combination of elements. They contain atoms that combine together to form molecules. In multicellular organisms, such as animals, molecules can interact to form cells that combine to form tissues, which make up organs. These combinations continue until entire multicellular organisms are formed.

All atoms contain protons, electrons, and neutrons ([link]). The only exception is hydrogen (H), which is made of one proton and one electron. A **proton** is a positively charged particle that resides in the **nucleus** (the core of the atom) of an atom and has a mass of 1 and a charge of +1. An **electron** is a negatively charged particle that travels in the space around the nucleus. In other words, it resides outside of the nucleus. It has a negligible mass and has a charge of –1.



Atoms are made up of protons and neutrons located within the nucleus, and electrons surrounding the nucleus.

Neutrons, like protons, reside in the nucleus of an atom. They have a mass of 1 and no charge. The positive (protons) and negative (electrons) charges balance each other in a neutral atom, which has a net zero charge.

Because protons and neutrons each have a mass of 1, the mass of an atom is equal to the number of protons and neutrons of that atom. The number of electrons does not factor into the overall mass, because their mass is so small.

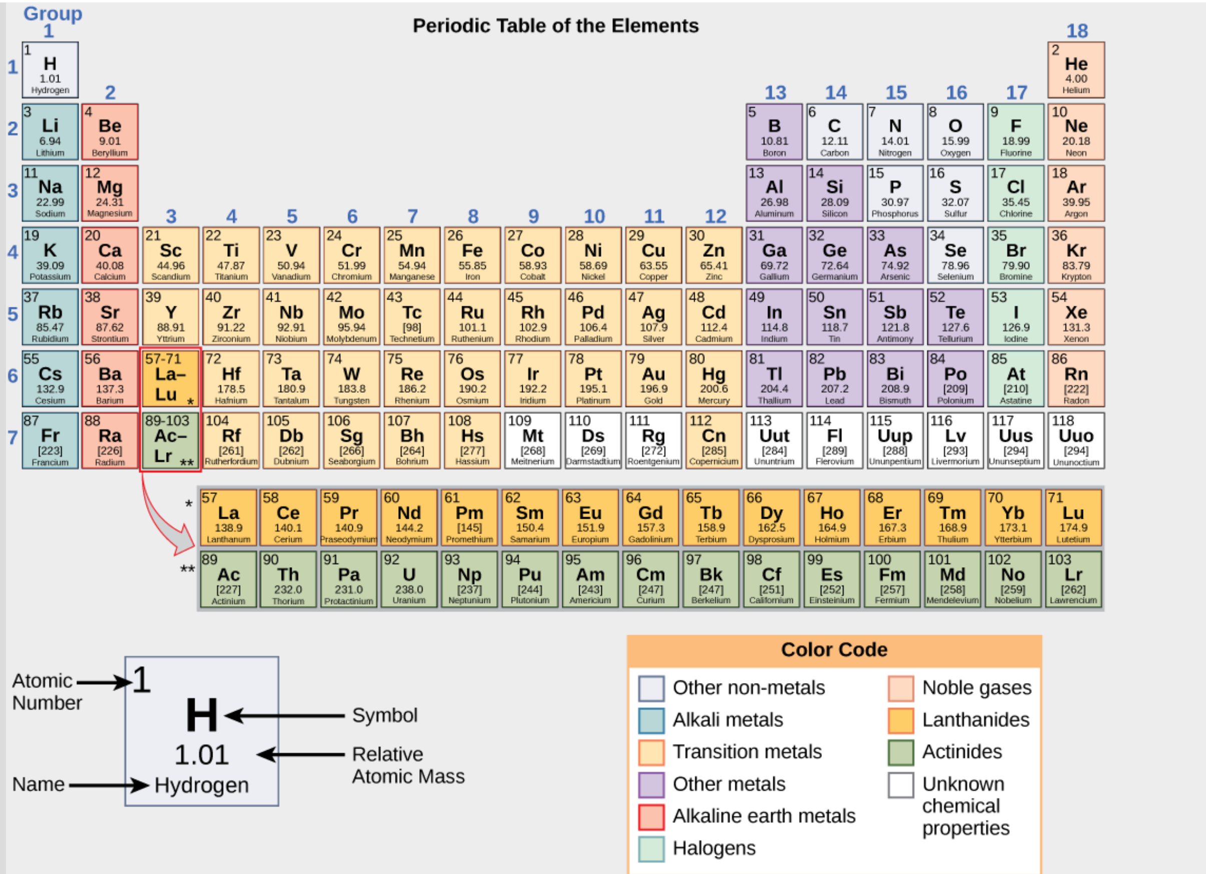
As stated earlier, each element has its own unique properties. Each contains a different number of protons and neutrons, giving it its own atomic number and mass number. The **atomic number** of an element is equal to the

number of protons that element contains. The **mass number**, or atomic mass, is the number of protons plus the number of neutrons of that element. Therefore, it is possible to determine the number of neutrons by subtracting the atomic number from the mass number.

These numbers provide information about the elements and how they will react when combined. Different elements have different melting and boiling points, and are in different states (liquid, solid, or gas) at room temperature. They also combine in different ways. Some form specific types of bonds, whereas others do not. How they combine is based on the number of electrons present. Because of these characteristics, the elements are arranged into the **periodic table of elements**, a chart of the elements that includes the atomic number and relative atomic mass of each element. The periodic table also provides key information about the properties of elements ([link])—often indicated by color-coding. The arrangement of the table also shows how the electrons in each element are organized and provides important details about how atoms will react with each other to form molecules.

Isotopes are different forms of the same element that have the same number of protons, but a different number of neutrons. Some elements, such as carbon, potassium, and uranium, have naturally occurring isotopes. Carbon-12, the most common isotope of carbon, contains six protons and six neutrons. Therefore, it has a mass number of 12 (six protons and six neutrons) and an atomic number of 6 (which makes it carbon). Carbon-14 contains six protons and eight neutrons. Therefore, it has a mass number of 14 (six protons and eight neutrons) and an atomic number of 6, meaning it is still the element carbon. These two alternate forms of carbon are isotopes. Some isotopes are unstable and will lose protons, other subatomic particles, or energy to form more stable elements. These are called **radioactive isotopes** or radioisotopes.

Note:
Art Connection



Arranged in columns and rows based on the characteristics of the elements, the periodic table provides key information about the elements and how they might interact with each other to form molecules. Most periodic tables provide a key or legend to the information they contain.

How many neutrons do (K) potassium-39 and potassium-40 have, respectively?

Note:

Evolution in Action

Carbon Dating

Carbon-14 (^{14}C) is a naturally occurring radioisotope that is created in the atmosphere by cosmic rays. This is a continuous process, so more ^{14}C is

always being created. As a living organism develops, the relative level of ^{14}C in its body is equal to the concentration of ^{14}C in the atmosphere. When an organism dies, it is no longer ingesting ^{14}C, so the ratio will decline. ^{14}C decays to ^{14}N by a process called beta decay; it gives off energy in this slow process.

After approximately 5,730 years, only one-half of the starting concentration of ^{14}C will have been converted to ^{14}N. The time it takes for half of the original concentration of an isotope to decay to its more stable form is called its half-life. Because the half-life of ^{14}C is long, it is used to age formerly living objects, such as fossils. Using the ratio of the ^{14}C concentration found in an object to the amount of ^{14}C detected in the atmosphere, the amount of the isotope that has not yet decayed can be determined. Based on this amount, the age of the fossil can be calculated to about 50,000 years ([link]). Isotopes with longer half-lives, such as potassium-40, are used to calculate the ages of older fossils. Through the use of carbon dating, scientists can reconstruct the ecology and biogeography of organisms living within the past 50,000 years.

The age of remains that contain carbon and are less than about 50,000 years old, such as this pygmy mammoth, can be determined using carbon dating. (credit: Bill Faulkner/NPS)

Note:
Concept in Action

To learn more about atoms and isotopes, and how you can tell one isotope from another, visit this site and run the simulation.

Chemical Bonds

How elements interact with one another depends on how their electrons are arranged and how many openings for electrons exist at the outermost region where electrons are present in an atom. Electrons exist at energy levels that form shells around the nucleus. The closest shell can hold up to two electrons. The closest shell to the nucleus is always filled first, before any other shell can be filled. Hydrogen has one electron; therefore, it has only one spot occupied within the lowest shell. Helium has two electrons; therefore, it can completely fill the lowest shell with its two electrons. If you look at the periodic table, you will see that hydrogen and helium are the only two elements in the first row. This is because they only have electrons in their first shell. Hydrogen and helium are the only two elements that have the lowest shell and no other shells.

The second and third energy levels can hold up to eight electrons. The eight electrons are arranged in four pairs and one position in each pair is filled with an electron before any pairs are completed.

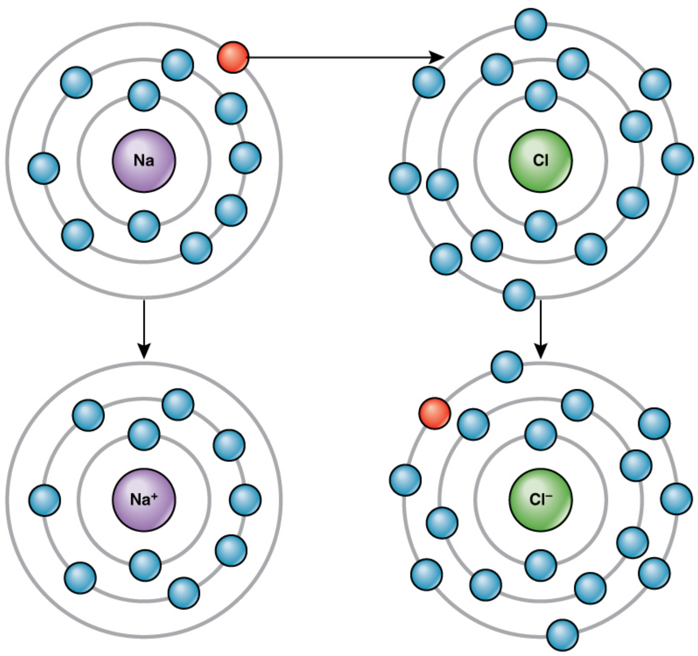
Looking at the periodic table again ([link]), you will notice that there are seven rows. These rows correspond to the number of shells that the elements within that row have. The elements within a particular row have increasing numbers of electrons as the columns proceed from left to right. Although each element has the same number of shells, not all of the shells are completely filled with electrons. If you look at the second row of the periodic table, you will find lithium (Li), beryllium (Be), boron (B), carbon (C), nitrogen (N), oxygen (O), fluorine (F), and neon (Ne). These all have electrons that occupy only the first and second shells. Lithium has only one electron in its outermost shell, beryllium has two electrons, boron has three, and so on, until the entire shell is filled with eight electrons, as is the case with neon.

Not all elements have enough electrons to fill their outermost shells, but an atom is at its most stable when all of the electron positions in the outermost shell are filled. Because of these vacancies in the outermost shells, we see the formation of **chemical bonds**, or interactions between two or more of the same or different elements that result in the formation of molecules. To achieve greater stability, atoms will tend to completely fill their outer shells and will bond with other elements to accomplish this goal by sharing electrons, accepting electrons from another atom, or donating electrons to another atom. Because the outermost shells of the elements with low atomic numbers (up to calcium, with atomic number 20) can hold eight electrons, this is referred to as the **octet rule**. An element can donate, accept, or share electrons with other elements to fill its outer shell and satisfy the octet rule.

When an atom does not contain equal numbers of protons and electrons, it is called an **ion**. Because the number of electrons does not equal the number of protons, each ion has a net charge. Positive ions are formed by losing electrons and are called **cations**. Negative ions are formed by gaining electrons and are called **anions**.

For example, sodium only has one electron in its outermost shell. It takes less energy for sodium to donate that one electron than it does to accept seven more electrons to fill the outer shell. If sodium loses an electron, it now has 11 protons and only 10 electrons, leaving it with an overall charge of +1. It is now called a sodium ion.

The chlorine atom has seven electrons in its outer shell. Again, it is more energy-efficient for chlorine to gain one electron than to lose seven. Therefore, it tends to gain an electron to create an ion with 17 protons and 18 electrons, giving it a net negative (–1) charge. It is now called a chloride ion. This movement of electrons from one element to another is referred to as **electron transfer**. As [link] illustrates, a sodium atom (Na) only has one electron in its outermost shell, whereas a chlorine atom (Cl) has seven electrons in its outermost shell. A sodium atom will donate its one electron to empty its shell, and a chlorine atom will accept that electron to fill its shell, becoming chloride. Both ions now satisfy the octet rule and have complete outermost shells. Because the number of electrons is no longer equal to the number of protons, each is now an ion and has a +1 (sodium) or –1 (chloride) charge.



Elements tend to fill their outermost shells with electrons. To do this, they can either donate or accept electrons from other elements.

Ionic Bonds

There are four types of bonds or interactions: ionic, covalent, hydrogen bonds, and van der Waals interactions. Ionic and covalent bonds are strong interactions that require a larger energy input to break apart. When an element donates an electron from its outer shell, as in the sodium atom example above, a positive ion is formed. The element accepting the electron is now negatively charged. Because positive and negative charges attract, these ions stay together and form an **ionic bond**, or a bond between ions. The elements bond together with the electron from one element staying predominantly with the other element. When Na^+ and Cl^- ions combine to produce NaCl, an electron from a sodium atom stays with the other seven from the chlorine atom, and the sodium and chloride ions attract each other in a lattice of ions with a net zero charge.

Covalent Bonds

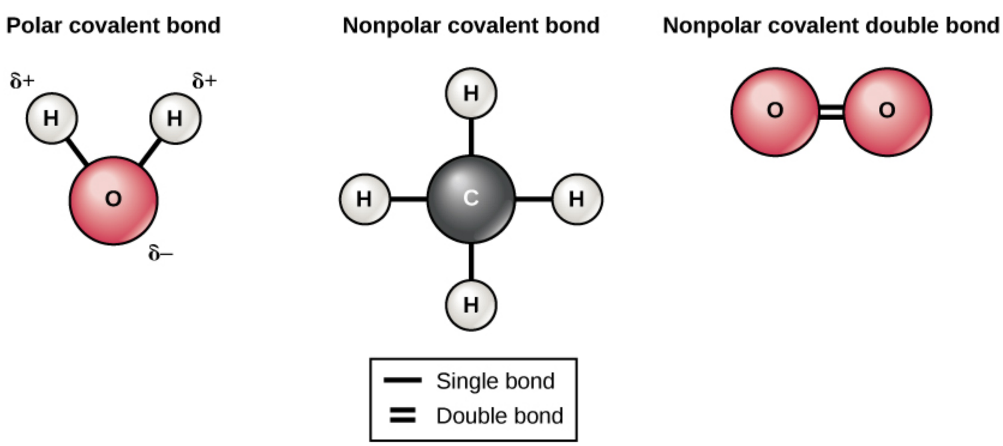
Another type of strong chemical bond between two or more atoms is a **covalent bond**. These bonds form when an electron is shared between two elements and are the strongest and most common form of chemical bond in living organisms. Covalent bonds form between the elements that make up the biological molecules in our cells. Unlike ionic bonds, covalent bonds do not dissociate in water.

The hydrogen and oxygen atoms that combine to form water molecules are bound together by covalent bonds. The electron from the hydrogen atom divides its time between the outer shell of the hydrogen atom and the incomplete outer shell of the oxygen atom. To completely fill the outer shell of an oxygen atom, two electrons from two hydrogen atoms are needed, hence the subscript “2” in H_2O. The electrons are shared between the atoms, dividing their time between them to “fill” the outer shell of each. This sharing is a lower energy state for all of the atoms involved than if they existed without their outer shells filled.

There are two types of covalent bonds: polar and nonpolar. **Nonpolar covalent bonds** form between two atoms of the same element or between

different elements that share the electrons equally. For example, an oxygen atom can bond with another oxygen atom to fill their outer shells. This association is nonpolar because the electrons will be equally distributed between each oxygen atom. Two covalent bonds form between the two oxygen atoms because oxygen requires two shared electrons to fill its outermost shell. Nitrogen atoms will form three covalent bonds (also called triple covalent) between two atoms of nitrogen because each nitrogen atom needs three electrons to fill its outermost shell. Another example of a nonpolar covalent bond is found in the methane (CH_4) molecule. The carbon atom has four electrons in its outermost shell and needs four more to fill it. It gets these four from four hydrogen atoms, each atom providing one. These elements all share the electrons equally, creating four nonpolar covalent bonds ([link]).

In a **polar covalent bond**, the electrons shared by the atoms spend more time closer to one nucleus than to the other nucleus. Because of the unequal distribution of electrons between the different nuclei, a slightly positive (δ+) or slightly negative (δ–) charge develops. The covalent bonds between hydrogen and oxygen atoms in water are polar covalent bonds. The shared electrons spend more time near the oxygen nucleus, giving it a small negative charge, than they spend near the hydrogen nuclei, giving these molecules a small positive charge.



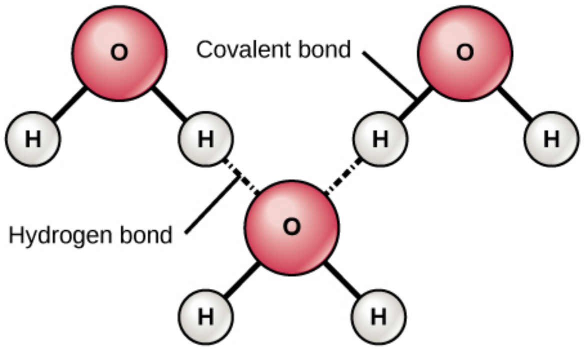
The water molecule (left) depicts a polar bond with a slightly positive charge on the hydrogen atoms and a

slightly negative charge on the oxygen. Examples of nonpolar bonds include methane (middle) and oxygen (right).

Hydrogen Bonds

Ionic and covalent bonds are strong bonds that require considerable energy to break. However, not all bonds between elements are ionic or covalent bonds. Weaker bonds can also form. These are attractions that occur between positive and negative charges that do not require much energy to break. Two weak bonds that occur frequently are hydrogen bonds and van der Waals interactions. These bonds give rise to the unique properties of water and the unique structures of DNA and proteins.

When polar covalent bonds containing a hydrogen atom form, the hydrogen atom in that bond has a slightly positive charge. This is because the shared electron is pulled more strongly toward the other element and away from the hydrogen nucleus. Because the hydrogen atom is slightly positive (δ+), it will be attracted to neighboring negative partial charges (δ–). When this happens, a weak interaction occurs between the δ+ charge of the hydrogen atom of one molecule and the δ– charge of the other molecule. This interaction is called a **hydrogen bond**. This type of bond is common; for example, the liquid nature of water is caused by the hydrogen bonds between water molecules ([link]). Hydrogen bonds give water the unique properties that sustain life. If it were not for hydrogen bonding, water would be a gas rather than a liquid at room temperature.



Hydrogen bonds form between slightly positive (δ+) and slightly negative (δ–) charges of polar covalent molecules, such as water.

Hydrogen bonds can form between different molecules and they do not always have to include a water molecule. Hydrogen atoms in polar bonds within any molecule can form bonds with other adjacent molecules. For example, hydrogen bonds hold together two long strands of DNA to give the DNA molecule its characteristic double-stranded structure. Hydrogen bonds are also responsible for some of the three-dimensional structure of proteins.

van der Waals Interactions

Like hydrogen bonds, **van der Waals interactions** are weak attractions or interactions between molecules. They occur between polar, covalently bound, atoms in different molecules. Some of these weak attractions are caused by temporary partial charges formed when electrons move around a nucleus. These weak interactions between molecules are important in biological systems.

Note:

Careers in Action

Radiography Technician

Have you or anyone you know ever had a magnetic resonance imaging (MRI) scan, a mammogram, or an X-ray? These tests produce images of your soft tissues and organs (as with an MRI or mammogram) or your bones (as happens in an X-ray) by using either radiowaves or special isotopes (radiolabeled or fluorescently labeled) that are ingested or injected into the body. These tests provide data for disease diagnoses by creating images of your organs or skeletal system.

MRI imaging works by subjecting hydrogen nuclei, which are abundant in the water in soft tissues, to fluctuating magnetic fields, which cause them to emit their own magnetic field. This signal is then read by sensors in the machine and interpreted by a computer to form a detailed image.

Some radiography technologists and technicians specialize in computed tomography, MRI, and mammography. They produce films or images of the body that help medical professionals examine and diagnose. Radiologists work directly with patients, explaining machinery, preparing them for exams, and ensuring that their body or body parts are positioned correctly to produce the needed images. Physicians or radiologists then analyze the test results.

Radiography technicians can work in hospitals, doctors' offices, or specialized imaging centers. Training to become a radiography technician happens at hospitals, colleges, and universities that offer certificates, associate's degrees, or bachelor's degrees in radiography.

Section Summary

Matter is anything that occupies space and has mass. It is made up of atoms of different elements. All of the 92 elements that occur naturally have unique qualities that allow them to combine in various ways to create compounds or molecules. Atoms, which consist of protons, neutrons, and electrons, are the smallest units of an element that retain all of the properties of that element. Electrons can be donated or shared between atoms to create bonds, including ionic, covalent, and hydrogen bonds, as well as van der Waals interactions.

Art Connections

Exercise:

Problem:

[link] How many neutrons do (K) potassium-39 and potassium-40 have, respectively?

Solution:

[link] Potassium-39 has twenty neutrons. Potassium-40 has twenty one neutrons.

Multiple Choice

Exercise:

Problem:

Magnesium has an atomic number of 12. Which of the following statements is true of a neutral magnesium atom?

a. It has 12 protons, 12 electrons, and 12 neutrons.
b. It has 12 protons, 12 electrons, and six neutrons.
c. It has six protons, six electrons, and no neutrons.
d. It has six protons, six electrons, and six neutrons.

Solution:

A

Exercise:

Problem: Which type of bond represents a weak chemical bond?

a. hydrogen bond
b. ionic bond

c. covalent bond
d. polar covalent bond

Solution:

A

Exercise:

Problem:

An isotope of sodium (Na) has a mass number of 22. How many neutrons does it have?

a. 11
b. 12
c. 22
d. 44

Solution:

A

Free Response

Exercise:

Problem:

Why are hydrogen bonds and van der Waals interactions necessary for cells?

Solution:

Hydrogen bonds and van der Waals interactions form weak associations between different molecules. They provide the structure and shape necessary for proteins and DNA within cells so that they

function properly. Hydrogen bonds also give water its unique properties, which are necessary for life.

Glossary

anion
: a negative ion formed by gaining electrons

atomic number
: the number of protons in an atom

cation
: a positive ion formed by losing electrons

chemical bond
: an interaction between two or more of the same or different elements that results in the formation of molecules

covalent bond
: a type of strong bond between two or more of the same or different elements; forms when electrons are shared between elements

electron
: a negatively charged particle that resides outside of the nucleus in the electron orbital; lacks functional mass and has a charge of –1

electron transfer
: the movement of electrons from one element to another

element
: one of 118 unique substances that cannot be broken down into smaller substances and retain the characteristic of that substance; each element has a specified number of protons and unique properties

hydrogen bond
: a weak bond between partially positively charged hydrogen atoms and partially negatively charged elements or molecules

ion

an atom or compound that does not contain equal numbers of protons and electrons, and therefore has a net charge

ionic bond

a chemical bond that forms between ions of opposite charges

isotope

one or more forms of an element that have different numbers of neutrons

mass number

the number of protons plus neutrons in an atom

matter

anything that has mass and occupies space

neutron

a particle with no charge that resides in the nucleus of an atom; has a mass of 1

nonpolar covalent bond

a type of covalent bond that forms between atoms when electrons are shared equally between atoms, resulting in no regions with partial charges as in polar covalent bonds

nucleus

(chemistry) the dense center of an atom made up of protons and (except in the case of a hydrogen atom) neutrons

octet rule

states that the outermost shell of an element with a low atomic number can hold eight electrons

periodic table of elements

an organizational chart of elements, indicating the atomic number and mass number of each element; also provides key information about the properties of elements

polar covalent bond

a type of covalent bond in which electrons are pulled toward one atom and away from another, resulting in slightly positive and slightly negative charged regions of the molecule

proton

a positively charged particle that resides in the nucleus of an atom; has a mass of 1 and a charge of +1

radioactive isotope

an isotope that spontaneously emits particles or energy to form a more stable element

van der Waals interaction

a weak attraction or interaction between molecules caused by slightly positively charged or slightly negatively charged atoms

Water
By the end of this section, you will be able to:

- Describe the properties of water that are critical to maintaining life

Do you ever wonder why scientists spend time looking for water on other planets? It is because water is essential to life; even minute traces of it on another planet can indicate that life could or did exist on that planet. Water is one of the more abundant molecules in living cells and the one most critical to life as we know it. Approximately 60–70 percent of your body is made up of water. Without it, life simply would not exist.

Water Is Polar

The hydrogen and oxygen atoms within water molecules form polar covalent bonds. The shared electrons spend more time associated with the oxygen atom than they do with hydrogen atoms. There is no overall charge to a water molecule, but there is a slight positive charge on each hydrogen atom and a slight negative charge on the oxygen atom. Because of these charges, the slightly positive hydrogen atoms repel each other and form the unique shape seen in [link]. Each water molecule attracts other water molecules because of the positive and negative charges in the different parts of the molecule. Water also attracts other polar molecules (such as sugars), forming hydrogen bonds. When a substance readily forms hydrogen bonds with water, it can dissolve in water and is referred to as **hydrophilic** ("water-loving"). Hydrogen bonds are not readily formed with nonpolar substances like oils and fats ([link]). These nonpolar compounds are **hydrophobic** ("water-fearing") and will not dissolve in water.

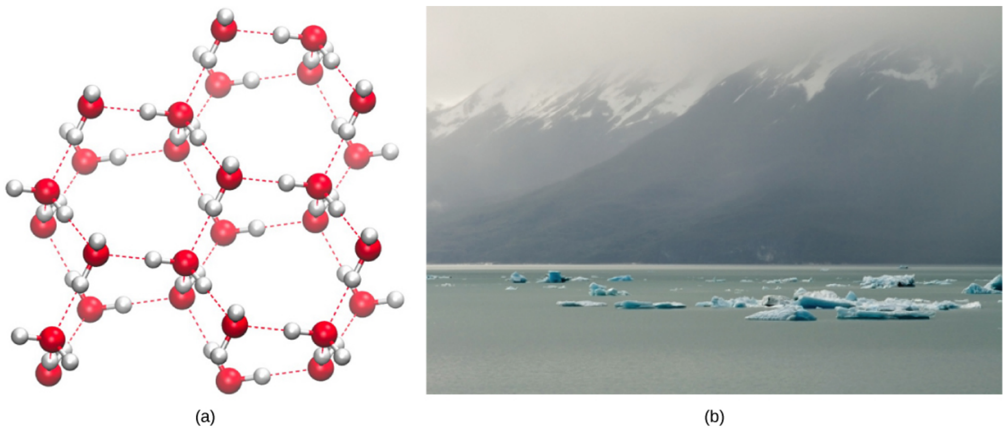
As this macroscopic image of oil and water show, oil is a nonpolar compound and, hence, will not dissolve in water. Oil and water do not mix. (credit: Gautam Dogra)

Water Stabilizes Temperature

The hydrogen bonds in water allow it to absorb and release heat energy more slowly than many other substances. **Temperature** is a measure of the motion (kinetic energy) of molecules. As the motion increases, energy is higher and thus temperature is higher. Water absorbs a great deal of energy before its temperature rises. Increased energy disrupts the hydrogen bonds between water molecules. Because these bonds can be created and disrupted rapidly, water absorbs an increase in energy and temperature changes only minimally. This means that water moderates temperature changes within organisms and in their environments. As energy input continues, the balance between hydrogen-bond formation and destruction swings toward the destruction side. More bonds are broken than are formed. This process results in the release of individual water molecules at the surface of the liquid (such as a body of water, the leaves of a plant, or the skin of an organism) in a process called **evaporation**. Evaporation of sweat, which is

90 percent water, allows for cooling of an organism, because breaking hydrogen bonds requires an input of energy and takes heat away from the body.

Conversely, as molecular motion decreases and temperatures drop, less energy is present to break the hydrogen bonds between water molecules. These bonds remain intact and begin to form a rigid, lattice-like structure (e.g., ice) ([link]**a**). When frozen, ice is less dense than liquid water (the molecules are farther apart). This means that ice floats on the surface of a body of water ([link]**b**). In lakes, ponds, and oceans, ice will form on the surface of the water, creating an insulating barrier to protect the animal and plant life beneath from freezing in the water. If this did not happen, plants and animals living in water would freeze in a block of ice and could not move freely, making life in cold temperatures difficult or impossible.



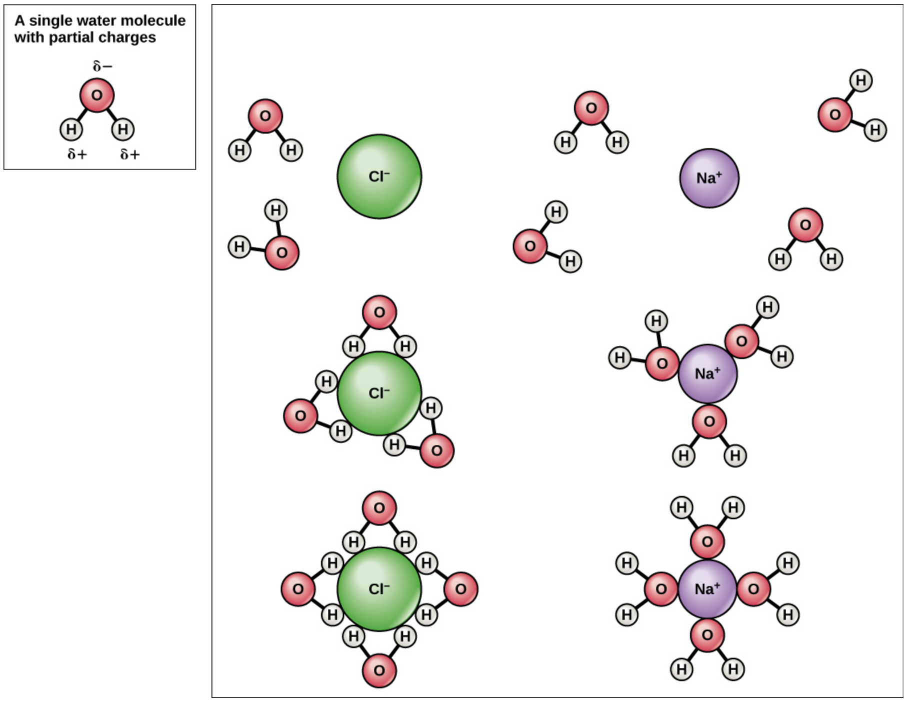
(a) The lattice structure of ice makes it less dense than the freely flowing molecules of liquid water. Ice's lower density enables it to (b) float on water. (credit a: modification of work by Jane Whitney; credit b: modification of work by Carlos Ponte)

Note:
Concepts in Action

Click here to see a 3-D animation of the structure of an ice lattice. (credit: image created by Jane Whitney using Visual Molecular Dynamics (VMD) software[footnote])
Humphrey, W., Dalke, A. and Schulten, K., "VMD—Visual Molecular Dynamics", *J. Molec. Graphics*, 1996, vol. 14, pp. 33-38. http://www.ks.uiuc.edu/Research/vmd/

Water Is an Excellent Solvent

Because water is polar, with slight positive and negative charges, ionic compounds and polar molecules can readily dissolve in it. Water is, therefore, what is referred to as a **solvent**—a substance capable of dissolving another substance. The charged particles will form hydrogen bonds with a surrounding layer of water molecules. This is referred to as a sphere of hydration and serves to keep the particles separated or dispersed in the water. In the case of table salt (NaCl) mixed in water ([link]), the sodium and chloride ions separate, or dissociate, in the water, and spheres of hydration are formed around the ions. A positively charged sodium ion is surrounded by the partially negative charges of oxygen atoms in water molecules. A negatively charged chloride ion is surrounded by the partially positive charges of hydrogen atoms in water molecules. These spheres of hydration are also referred to as hydration shells. The polarity of the water molecule makes it an effective solvent and is important in its many roles in living systems.

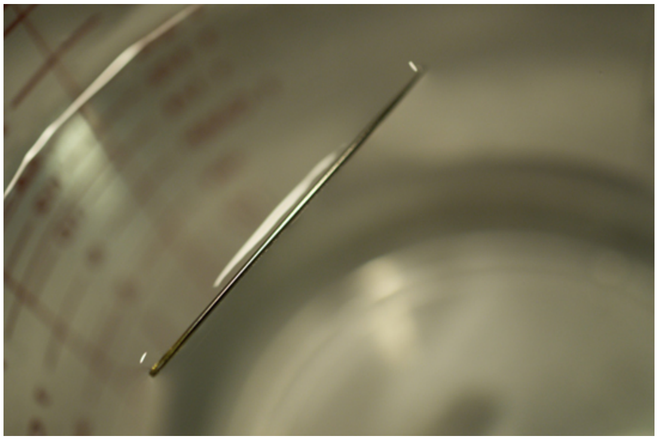


When table salt (NaCl) is mixed in water, spheres of hydration form around the ions.

Water Is Cohesive

Have you ever filled up a glass of water to the very top and then slowly added a few more drops? Before it overflows, the water actually forms a dome-like shape above the rim of the glass. This water can stay above the glass because of the property of **cohesion**. In cohesion, water molecules are attracted to each other (because of hydrogen bonding), keeping the molecules together at the liquid-air (gas) interface, although there is no more room in the glass. Cohesion gives rise to **surface tension**, the capacity of a substance to withstand rupture when placed under tension or stress. When you drop a small scrap of paper onto a droplet of water, the paper floats on top of the water droplet, although the object is denser (heavier) than the water. This occurs because of the surface tension that is created by the water molecules. Cohesion and surface tension keep the water molecules intact and the item floating on the top. It is even possible to

“float” a steel needle on top of a glass of water if you place it gently, without breaking the surface tension ([link]).



The weight of a needle on top of water pulls the surface tension downward; at the same time, the surface tension of the water is pulling it up, suspending the needle on the surface of the water and keeping it from sinking. Notice the indentation in the water around the needle. (credit: Cory Zanker)

These cohesive forces are also related to the water’s property of **adhesion**, or the attraction between water molecules and other molecules. This is observed when water “climbs” up a straw placed in a glass of water. You will notice that the water appears to be higher on the sides of the straw than in the middle. This is because the water molecules are attracted to the straw and therefore adhere to it.

Cohesive and adhesive forces are important for sustaining life. For example, because of these forces, water can flow up from the roots to the tops of plants to feed the plant.

Note:
Concept in Action

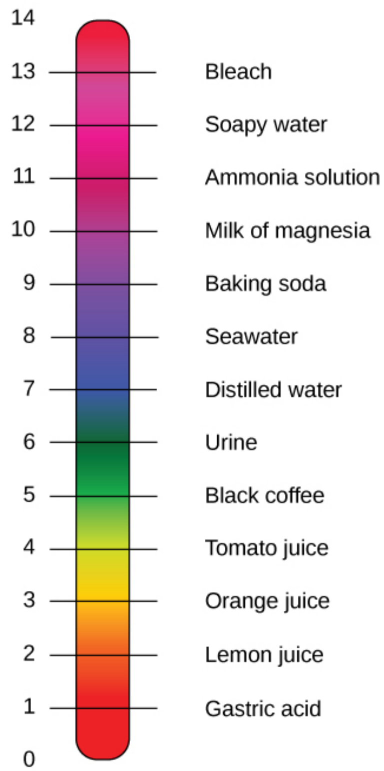
To learn more about water, visit the U.S. Geological Survey Water Science for Schools: All About Water! website.

Buffers, pH, Acids, and Bases

The pH of a solution is a measure of its acidity or alkalinity. You have probably used **litmus paper**, paper that has been treated with a natural water-soluble dye so it can be used as a pH indicator, to test how much acid or base (alkalinity) exists in a solution. You might have even used some to make sure the water in an outdoor swimming pool is properly treated. In both cases, this pH test measures the amount of hydrogen ions that exists in a given solution. High concentrations of hydrogen ions yield a low pH, whereas low levels of hydrogen ions result in a high pH. The overall concentration of hydrogen ions is inversely related to its pH and can be measured on the **pH scale** ([link]). Therefore, the more hydrogen ions present, the lower the pH; conversely, the fewer hydrogen ions, the higher the pH.

The pH scale ranges from 0 to 14. A change of one unit on the pH scale represents a change in the concentration of hydrogen ions by a factor of 10, a change in two units represents a change in the concentration of hydrogen ions by a factor of 100. Thus, small changes in pH represent large changes in the concentrations of hydrogen ions. Pure water is neutral. It is neither acidic nor basic, and has a pH of 7.0. Anything below 7.0 (ranging from 0.0 to 6.9) is acidic, and anything above 7.0 (from 7.1 to 14.0) is alkaline. The blood in your veins is slightly alkaline (pH = 7.4). The environment in your

stomach is highly acidic (pH = 1 to 2). Orange juice is mildly acidic (pH = approximately 3.5), whereas baking soda is basic (pH = 9.0).



The pH scale measures the amount of hydrogen ions (H^+) in a substance. (credit: modification of work by Edward Stevens)

Acids are substances that provide hydrogen ions (H^+) and lower pH, whereas **bases** provide hydroxide ions (OH^-) and raise pH. The stronger the acid, the more readily it donates H^+. For example, hydrochloric acid and lemon juice are very acidic and readily give up H^+ when added to water. Conversely, bases are those substances that readily donate OH^-. The OH^- ions combine with H^+ to produce water, which raises a substance's pH. Sodium hydroxide and many household cleaners are very alkaline and give up OH^- rapidly when placed in water, thereby raising the pH.

Most cells in our bodies operate within a very narrow window of the pH scale, typically ranging only from 7.2 to 7.6. If the pH of the body is outside of this range, the respiratory system malfunctions, as do other organs in the body. Cells no longer function properly, and proteins will break down. Deviation outside of the pH range can induce coma or even cause death.

So how is it that we can ingest or inhale acidic or basic substances and not die? Buffers are the key. **Buffers** readily absorb excess H^+ or $OH^–$, keeping the pH of the body carefully maintained in the aforementioned narrow range. Carbon dioxide is part of a prominent buffer system in the human body; it keeps the pH within the proper range. This buffer system involves carbonic acid (H_2CO_3) and bicarbonate ($HCO_3^–$) anion. If too much H^+ enters the body, bicarbonate will combine with the H^+ to create carbonic acid and limit the decrease in pH. Likewise, if too much $OH^–$ is introduced into the system, carbonic acid will rapidly dissociate into bicarbonate and H^+ ions. The H^+ ions can combine with the $OH^–$ ions, limiting the increase in pH. While carbonic acid is an important product in this reaction, its presence is fleeting because the carbonic acid is released from the body as carbon dioxide gas each time we breathe. Without this buffer system, the pH in our bodies would fluctuate too much and we would fail to survive.

Section Summary

Water has many properties that are critical to maintaining life. It is polar, allowing for the formation of hydrogen bonds, which allow ions and other polar molecules to dissolve in water. Therefore, water is an excellent solvent. The hydrogen bonds between water molecules give water the ability to hold heat better than many other substances. As the temperature rises, the hydrogen bonds between water continually break and reform, allowing for the overall temperature to remain stable, although increased energy is added to the system. Water’s cohesive forces allow for the property of surface tension. All of these unique properties of water are important in the chemistry of living organisms.

The pH of a solution is a measure of the concentration of hydrogen ions in the solution. A solution with a high number of hydrogen ions is acidic and has a low pH value. A solution with a high number of hydroxide ions is

basic and has a high pH value. The pH scale ranges from 0 to 14, with a pH of 7 being neutral. Buffers are solutions that moderate pH changes when an acid or base is added to the buffer system. Buffers are important in biological systems because of their ability to maintain constant pH conditions.

Multiple Choice

Exercise:

Problem: Which of the following statements is not true?

a. Water is polar.
b. Water stabilizes temperature.
c. Water is essential for life.
d. Water is the most abundant atom in Earth’s atmosphere.

Solution:

D

Exercise:

Problem:

Using a pH meter, you find the pH of an unknown solution to be 8.0. How would you describe this solution?

a. weakly acidic
b. strongly acidic
c. weakly basic
d. strongly basic

Solution:

C

Exercise:

Problem:

The pH of lemon juice is about 2.0, whereas tomato juice's pH is about 4.0. Approximately how much of an increase in hydrogen ion concentration is there between tomato juice and lemon juice?

a. 2 times
b. 10 times
c. 100 times
d. 1000 times

Solution:

C

Free Response

Exercise:

Problem: Why can some insects walk on water?

Solution:

Some insects can walk on water, although they are heavier (denser) than water, because of the surface tension of water. Surface tension results from cohesion, or the attraction between water molecules at the surface of the body of water [the liquid-air (gas) interface].

Exercise:

Problem: Explain why water is an excellent solvent.

Solution:

Water molecules are polar, meaning they have separated partial positive and negative charges. Because of these charges, water molecules are able to surround charged particles created when a substance dissociates. The surrounding layer of water molecules stabilizes the ion and keeps differently charged ions from reassociating, so the substance stays dissolved.

Glossary

acid
: a substance that donates hydrogen ions and therefore lowers pH

adhesion
: the attraction between water molecules and molecules of a different substance

base
: a substance that absorbs hydrogen ions and therefore raises pH

buffer
: a solution that resists a change in pH by absorbing or releasing hydrogen or hydroxide ions

cohesion
: the intermolecular forces between water molecules caused by the polar nature of water; creates surface tension

evaporation
: the release of water molecules from liquid water to form water vapor

hydrophilic
: describes a substance that dissolves in water; water-loving

hydrophobic
: describes a substance that does not dissolve in water; water-fearing

litmus paper

filter paper that has been treated with a natural water-soluble dye so it can be used as a pH indicator

pH scale
: a scale ranging from 0 to 14 that measures the approximate concentration of hydrogen ions of a substance

solvent
: a substance capable of dissolving another substance

surface tension
: the cohesive force at the surface of a body of liquid that prevents the molecules from separating

temperature
: a measure of molecular motion

Biological Molecules
By the end of this section, you will be able to:

- Describe the ways in which carbon is critical to life
- Explain the impact of slight changes in amino acids on organisms
- Describe the four major types of biological molecules
- Understand the functions of the four major types of molecules

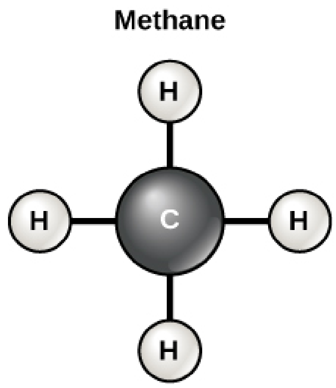
The large molecules necessary for life that are built from smaller organic molecules are called biological **macromolecules**. There are four major classes of biological macromolecules (carbohydrates, lipids, proteins, and nucleic acids), and each is an important component of the cell and performs a wide array of functions. Combined, these molecules make up the majority of a cell's mass. Biological macromolecules are organic, meaning that they contain carbon (with some exceptions, like carbon dioxide). In addition, they may contain hydrogen, oxygen, nitrogen, phosphorus, sulfur, and additional minor elements.

Carbon

It is often said that life is "carbon-based." This means that carbon atoms, bonded to other carbon atoms or other elements, form the fundamental components of many, if not most, of the molecules found uniquely in living things. Other elements play important roles in biological molecules, but carbon certainly qualifies as the "foundation" element for molecules in living things. It is the bonding properties of carbon atoms that are responsible for its important role.

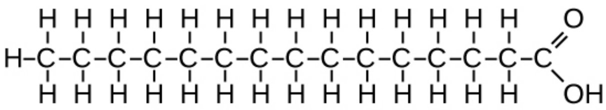
Carbon Bonding

Carbon contains four electrons in its outer shell. Therefore, it can form four covalent bonds with other atoms or molecules. The simplest organic carbon molecule is methane (CH_4), in which four hydrogen atoms bind to a carbon atom ([link]).

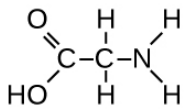


Carbon can form four covalent bonds to create an organic molecule. The simplest carbon molecule is methane (CH_4), depicted here.

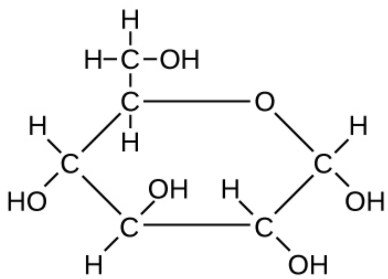
However, structures that are more complex are made using carbon. Any of the hydrogen atoms can be replaced with another carbon atom covalently bonded to the first carbon atom. In this way, long and branching chains of carbon compounds can be made ([link]**a**). The carbon atoms may bond with atoms of other elements, such as nitrogen, oxygen, and phosphorus ([link]**b**). The molecules may also form rings, which themselves can link with other rings ([link]**c**). This diversity of molecular forms accounts for the diversity of functions of the biological macromolecules and is based to a large degree on the ability of carbon to form multiple bonds with itself and other atoms.



(a)



(b)



(c)

These examples show three molecules (found in living organisms) that contain carbon atoms bonded in various ways to other carbon atoms and the atoms of other elements. (a) This molecule of stearic acid has a long chain of carbon atoms. (b) Glycine, a component of proteins, contains carbon, nitrogen, oxygen, and hydrogen atoms. (c) Glucose, a sugar, has a ring of carbon atoms and one oxygen atom.

Carbohydrates

Carbohydrates are macromolecules with which most consumers are somewhat familiar. To lose weight, some individuals adhere to "low-carb" diets. Athletes, in contrast, often "carb-load" before important competitions to ensure that they have sufficient energy to compete at a high level. Carbohydrates are, in fact, an essential part of our diet; grains, fruits, and vegetables are all natural sources of carbohydrates. Carbohydrates provide energy to the body, particularly through glucose, a simple sugar. Carbohydrates also have other important functions in humans, animals, and plants.

Carbohydrates can be represented by the formula $(CH_2O)_n$, where n is the number of carbon atoms in the molecule. In other words, the ratio of carbon to hydrogen to oxygen is 1:2:1 in carbohydrate molecules. Carbohydrates are classified into three subtypes: monosaccharides, disaccharides, and polysaccharides.

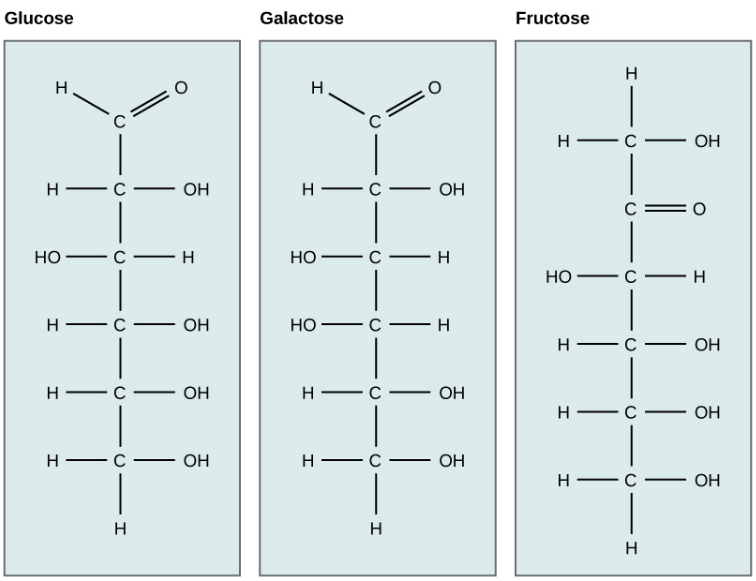
Monosaccharides (mono- = "one"; sacchar- = "sweet") are simple sugars, the most common of which is glucose. In monosaccharides, the number of carbon atoms usually ranges from three to six. Most monosaccharide names end with the suffix -ose. Depending on the number of carbon atoms in the sugar, they may be known as trioses (three carbon atoms), pentoses (five carbon atoms), and hexoses (six carbon atoms).

Monosaccharides may exist as a linear chain or as ring-shaped molecules; in aqueous solutions, they are usually found in the ring form.

The chemical formula for glucose is $C_6H_{12}O_6$. In most living species, glucose is an important source of energy. During cellular respiration, energy is released from glucose, and that energy is used to help make adenosine triphosphate (ATP). Plants synthesize glucose using carbon dioxide and water by the process of photosynthesis, and the glucose, in turn, is used for the energy requirements of the plant. The excess synthesized glucose is often stored as starch that is broken down by other organisms that feed on plants.

Galactose (part of lactose, or milk sugar) and fructose (found in fruit) are other common monosaccharides. Although glucose, galactose, and fructose all have the same chemical formula ($C_6H_{12}O_6$), they differ structurally and

chemically (and are known as isomers) because of differing arrangements of atoms in the carbon chain ([link]).



Glucose, galactose, and fructose are isomeric monosaccharides, meaning that they have the same chemical formula but slightly different structures.

Disaccharides (di- = “two”) form when two monosaccharides undergo a dehydration reaction (a reaction in which the removal of a water molecule occurs). During this process, the hydroxyl group (–OH) of one monosaccharide combines with a hydrogen atom of another monosaccharide, releasing a molecule of water (H_2O) and forming a covalent bond between atoms in the two sugar molecules.

Common disaccharides include lactose, maltose, and sucrose. Lactose is a disaccharide consisting of the monomers glucose and galactose. It is found naturally in milk. Maltose, or malt sugar, is a disaccharide formed from a dehydration reaction between two glucose molecules. The most common

disaccharide is sucrose, or table sugar, which is composed of the monomers glucose and fructose.

A long chain of monosaccharides linked by covalent bonds is known as a **polysaccharide** (poly- = “many”). The chain may be branched or unbranched, and it may contain different types of monosaccharides. Polysaccharides may be very large molecules. Starch, glycogen, cellulose, and chitin are examples of polysaccharides.

Starch is the stored form of sugars in plants and is made up of amylose and amylopectin (both polymers of glucose). Plants are able to synthesize glucose, and the excess glucose is stored as starch in different plant parts, including roots and seeds. The starch that is consumed by animals is broken down into smaller molecules, such as glucose. The cells can then absorb the glucose.

Glycogen is the storage form of glucose in humans and other vertebrates, and is made up of monomers of glucose. Glycogen is the animal equivalent of starch and is a highly branched molecule usually stored in liver and muscle cells. Whenever glucose levels decrease, glycogen is broken down to release glucose.

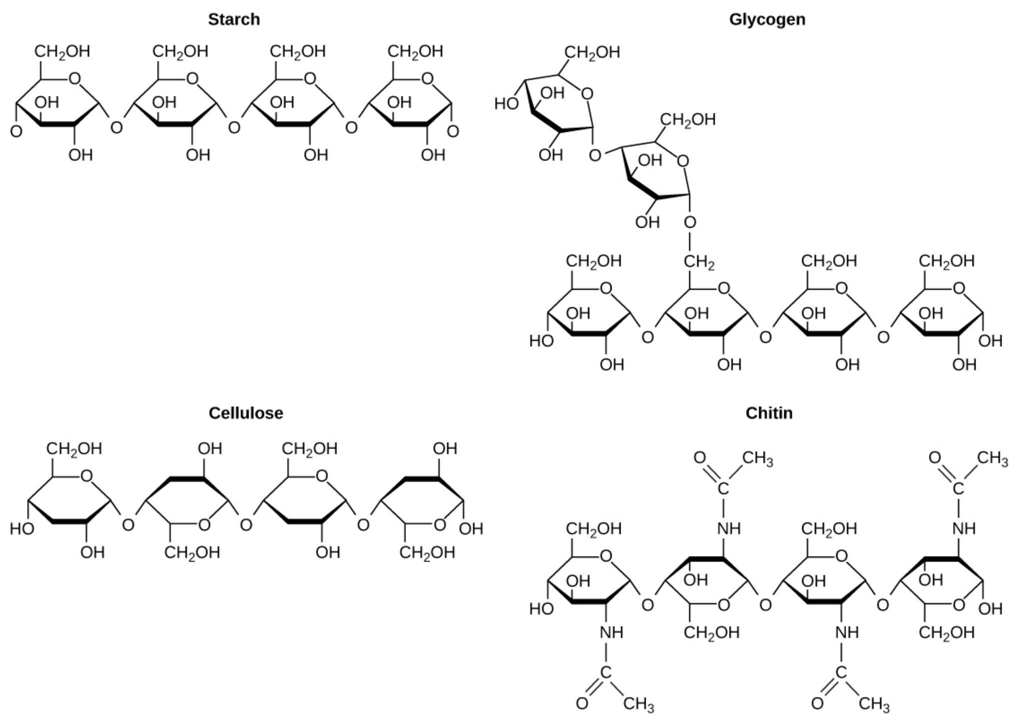
Cellulose is one of the most abundant natural biopolymers. The cell walls of plants are mostly made of cellulose, which provides structural support to the cell. Wood and paper are mostly cellulosic in nature. Cellulose is made up of glucose monomers that are linked by bonds between particular carbon atoms in the glucose molecule.

Every other glucose monomer in cellulose is flipped over and packed tightly as extended long chains. This gives cellulose its rigidity and high tensile strength—which is so important to plant cells. Cellulose passing through our digestive system is called dietary fiber. While the glucose-glucose bonds in cellulose cannot be broken down by human digestive enzymes, herbivores such as cows, buffalos, and horses are able to digest grass that is rich in cellulose and use it as a food source. In these animals, certain species of bacteria reside in the rumen (part of the digestive system of herbivores) and secrete the enzyme cellulase. The appendix also contains bacteria that break down cellulose, giving it an important role in the

digestive systems of ruminants. Cellulases can break down cellulose into glucose monomers that can be used as an energy source by the animal.

Carbohydrates serve other functions in different animals. Arthropods, such as insects, spiders, and crabs, have an outer skeleton, called the exoskeleton, which protects their internal body parts. This exoskeleton is made of the biological macromolecule **chitin**, which is a nitrogenous carbohydrate. It is made of repeating units of a modified sugar containing nitrogen.

Thus, through differences in molecular structure, carbohydrates are able to serve the very different functions of energy storage (starch and glycogen) and structural support and protection (cellulose and chitin) ([link]).



Although their structures and functions differ, all polysaccharide carbohydrates are made up of monosaccharides and have the chemical formula $(CH_2O)n$.

Note:
Careers in Action
Registered Dietitian
Obesity is a worldwide health concern, and many diseases, such as diabetes and heart disease, are becoming more prevalent because of obesity. This is one of the reasons why registered dietitians are increasingly sought after for advice. Registered dietitians help plan food and nutrition programs for individuals in various settings. They often work with patients in health-care facilities, designing nutrition plans to prevent and treat diseases. For example, dietitians may teach a patient with diabetes how to manage blood-sugar levels by eating the correct types and amounts of carbohydrates. Dietitians may also work in nursing homes, schools, and private practices.
To become a registered dietitian, one needs to earn at least a bachelor's degree in dietetics, nutrition, food technology, or a related field. In addition, registered dietitians must complete a supervised internship program and pass a national exam. Those who pursue careers in dietetics take courses in nutrition, chemistry, biochemistry, biology, microbiology, and human physiology. Dietitians must become experts in the chemistry and functions of food (proteins, carbohydrates, and fats).

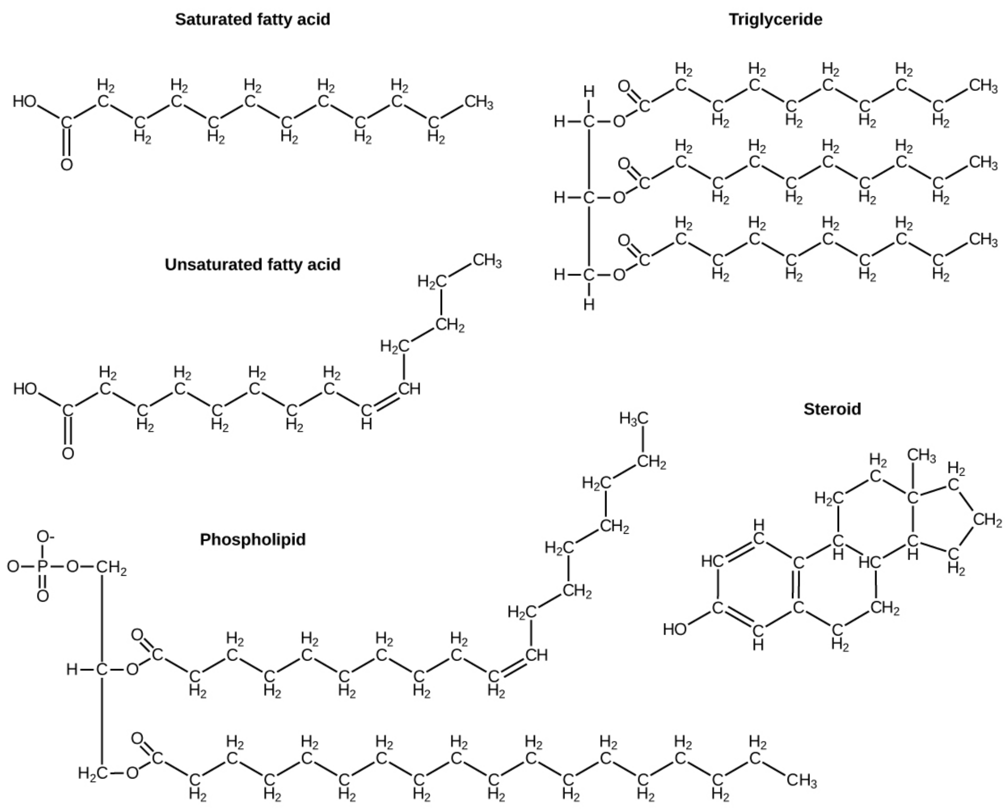
Lipids

Lipids include a diverse group of compounds that are united by a common feature. **Lipids** are hydrophobic ("water-fearing"), or insoluble in water, because they are nonpolar molecules. This is because they are hydrocarbons that include only nonpolar carbon-carbon or carbon-hydrogen bonds. Lipids perform many different functions in a cell. Cells store energy for long-term use in the form of lipids called fats. Lipids also provide insulation from the environment for plants and animals ([link]). For example, they help keep aquatic birds and mammals dry because of their water-repelling nature. Lipids are also the building blocks of many hormones and are an important

constituent of the plasma membrane. Lipids include fats, oils, waxes, phospholipids, and steroids.

Hydrophobic lipids in the fur of aquatic mammals, such as this river otter, protect them from the elements. (credit: Ken Bosma)

A **fat** molecule, such as a triglyceride, consists of two main components—glycerol and fatty acids. Glycerol is an organic compound with three carbon atoms, five hydrogen atoms, and three hydroxyl (–OH) groups. Fatty acids have a long chain of hydrocarbons to which an acidic carboxyl group is attached, hence the name “fatty acid.” The number of carbons in the fatty acid may range from 4 to 36; most common are those containing 12–18 carbons. In a fat molecule, a fatty acid is attached to each of the three oxygen atoms in the –OH groups of the glycerol molecule with a covalent bond ([link]).



Lipids include fats, such as triglycerides, which are made up of fatty acids and glycerol, phospholipids, and steroids.

During this covalent bond formation, three water molecules are released. The three fatty acids in the fat may be similar or dissimilar. These fats are also called **triglycerides** because they have three fatty acids. Some fatty acids have common names that specify their origin. For example, palmitic acid, a saturated fatty acid, is derived from the palm tree. Arachidic acid is derived from *Arachis hypogaea*, the scientific name for peanuts.

Fatty acids may be saturated or unsaturated. In a fatty acid chain, if there are only single bonds between neighboring carbons in the hydrocarbon chain, the fatty acid is saturated. **Saturated fatty acids** are saturated with hydrogen; in other words, the number of hydrogen atoms attached to the carbon skeleton is maximized.

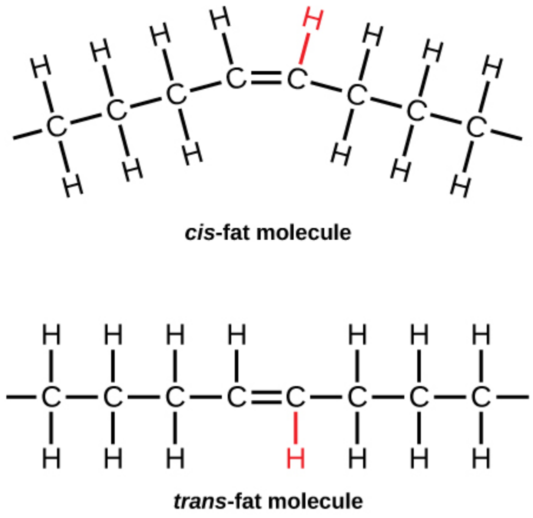
When the hydrocarbon chain contains a double bond, the fatty acid is an **unsaturated fatty acid**.

Most unsaturated fats are liquid at room temperature and are called **oils**. If there is one double bond in the molecule, then it is known as a monounsaturated fat (e.g., olive oil), and if there is more than one double bond, then it is known as a polyunsaturated fat (e.g., canola oil).

Saturated fats tend to get packed tightly and are solid at room temperature. Animal fats with stearic acid and palmitic acid contained in meat, and the fat with butyric acid contained in butter, are examples of saturated fats. Mammals store fats in specialized cells called adipocytes, where globules of fat occupy most of the cell. In plants, fat or oil is stored in seeds and is used as a source of energy during embryonic development.

Unsaturated fats or oils are usually of plant origin and contain unsaturated fatty acids. The double bond causes a bend or a “kink” that prevents the fatty acids from packing tightly, keeping them liquid at room temperature. Olive oil, corn oil, canola oil, and cod liver oil are examples of unsaturated fats. Unsaturated fats help to improve blood cholesterol levels, whereas saturated fats contribute to plaque formation in the arteries, which increases the risk of a heart attack.

In the food industry, oils are artificially hydrogenated to make them semi-solid, leading to less spoilage and increased shelf life. Simply speaking, hydrogen gas is bubbled through oils to solidify them. During this hydrogenation process, double bonds of the *cis*-conformation in the hydrocarbon chain may be converted to double bonds in the *trans*-conformation. This forms a ***trans*-fat** from a *cis*-fat. The orientation of the double bonds affects the chemical properties of the fat ([link]).



During the hydrogenation process, the orientation around the double bonds is changed, making a *trans*-fat from a *cis*-fat. This changes the chemical properties of the molecule.

Margarine, some types of peanut butter, and shortening are examples of artificially hydrogenated *trans*-fats. Recent studies have shown that an increase in *trans*-fats in the human diet may lead to an increase in levels of low-density lipoprotein (LDL), or “bad” cholesterol, which, in turn, may lead to plaque deposition in the arteries, resulting in heart disease. Many fast food restaurants have recently eliminated the use of *trans*-fats, and U.S. food labels are now required to list their *trans*-fat content.

Essential fatty acids are fatty acids that are required but not synthesized by the human body. Consequently, they must be supplemented through the diet. Omega-3 fatty acids fall into this category and are one of only two known essential fatty acids for humans (the other being omega-6 fatty acids). They are a type of polyunsaturated fat and are called omega-3 fatty acids because the third carbon from the end of the fatty acid participates in a double bond.

Salmon, trout, and tuna are good sources of omega-3 fatty acids. Omega-3 fatty acids are important in brain function and normal growth and development. They may also prevent heart disease and reduce the risk of cancer.

Like carbohydrates, fats have received a lot of bad publicity. It is true that eating an excess of fried foods and other “fatty” foods leads to weight gain. However, fats do have important functions. Fats serve as long-term energy storage. They also provide insulation for the body. Therefore, “healthy” unsaturated fats in moderate amounts should be consumed on a regular basis.

Phospholipids are the major constituent of the plasma membrane. Like fats, they are composed of fatty acid chains attached to a glycerol or similar backbone. Instead of three fatty acids attached, however, there are two fatty acids and the third carbon of the glycerol backbone is bound to a phosphate group. The phosphate group is modified by the addition of an alcohol.

A phospholipid has both hydrophobic and hydrophilic regions. The fatty acid chains are hydrophobic and exclude themselves from water, whereas the phosphate is hydrophilic and interacts with water.

Cells are surrounded by a membrane, which has a bilayer of phospholipids. The fatty acids of phospholipids face inside, away from water, whereas the phosphate group can face either the outside environment or the inside of the cell, which are both aqueous.

Steroids and Waxes

Unlike the phospholipids and fats discussed earlier, **steroids** have a ring structure. Although they do not resemble other lipids, they are grouped with them because they are also hydrophobic. All steroids have four, linked carbon rings and several of them, like cholesterol, have a short tail.

Cholesterol is a steroid. Cholesterol is mainly synthesized in the liver and is the precursor of many steroid hormones, such as testosterone and estradiol.

It is also the precursor of vitamins E and K. Cholesterol is the precursor of bile salts, which help in the breakdown of fats and their subsequent absorption by cells. Although cholesterol is often spoken of in negative terms, it is necessary for the proper functioning of the body. It is a key component of the plasma membranes of animal cells.

Waxes are made up of a hydrocarbon chain with an alcohol (–OH) group and a fatty acid. Examples of animal waxes include beeswax and lanolin. Plants also have waxes, such as the coating on their leaves, that helps prevent them from drying out.

Note:
Concept in Action

For an additional perspective on lipids, explore “Biomolecules: The Lipids” through this interactive animation.

Proteins

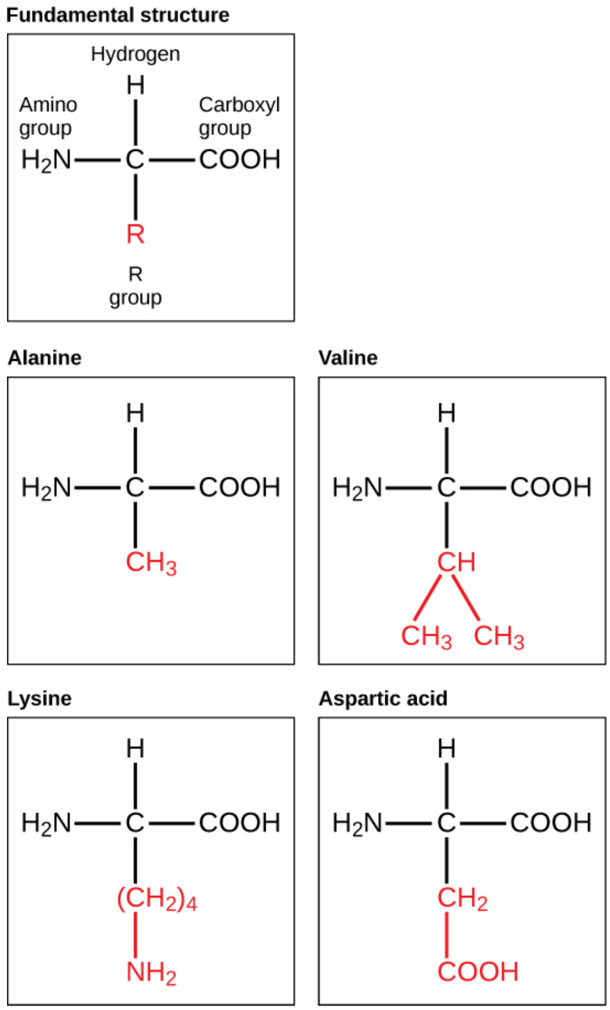
Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. Proteins may be structural, regulatory, contractile, or protective; they may serve in transport, storage, or membranes; or they may be toxins or enzymes. Each cell in a living system may contain thousands of different proteins, each with a unique function. Their structures, like their functions, vary greatly. They are all, however, polymers of amino acids, arranged in a linear sequence.

The functions of proteins are very diverse because there are 20 different chemically distinct amino acids that form long chains, and the amino acids can be in any order. For example, proteins can function as enzymes or hormones. **Enzymes**, which are produced by living cells, are catalysts in biochemical reactions (like digestion) and are usually proteins. Each enzyme is specific for the substrate (a reactant that binds to an enzyme) upon which it acts. Enzymes can function to break molecular bonds, to rearrange bonds, or to form new bonds. An example of an enzyme is salivary amylase, which breaks down amylose, a component of starch.

Hormones are chemical signaling molecules, usually proteins or steroids, secreted by an endocrine gland or group of endocrine cells that act to control or regulate specific physiological processes, including growth, development, metabolism, and reproduction. For example, insulin is a protein hormone that maintains blood glucose levels.

Proteins have different shapes and molecular weights; some proteins are globular in shape whereas others are fibrous in nature. For example, hemoglobin is a globular protein, but collagen, found in our skin, is a fibrous protein. Protein shape is critical to its function. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to a loss of function or **denaturation** (to be discussed in more detail later). All proteins are made up of different arrangements of the same 20 kinds of amino acids.

Amino acids are the monomers that make up proteins. Each amino acid has the same fundamental structure, which consists of a central carbon atom bonded to an amino group ($–NH_2$), a carboxyl group (–COOH), and a hydrogen atom. Every amino acid also has another variable atom or group of atoms bonded to the central carbon atom known as the R group. The R group is the only difference in structure between the 20 amino acids; otherwise, the amino acids are identical ([link]).



Amino acids are made up of a central carbon bonded to an amino group ($–NH_2$), a carboxyl group (–COOH), and a hydrogen atom. The central carbon's fourth bond varies among the different amino acids, as seen in these examples of alanine, valine, lysine, and aspartic acid.

The chemical nature of the R group determines the chemical nature of the amino acid within its protein (that is, whether it is acidic, basic, polar, or nonpolar).

The sequence and number of amino acids ultimately determine a protein’s shape, size, and function. Each amino acid is attached to another amino acid by a covalent bond, known as a peptide bond, which is formed by a dehydration reaction. The carboxyl group of one amino acid and the amino group of a second amino acid combine, releasing a water molecule. The resulting bond is the peptide bond.

The products formed by such a linkage are called polypeptides. While the terms polypeptide and protein are sometimes used interchangeably, a **polypeptide** is technically a polymer of amino acids, whereas the term protein is used for a polypeptide or polypeptides that have combined together, have a distinct shape, and have a unique function.

Note:
Evolution in Action
The Evolutionary Significance of Cytochrome c
Cytochrome c is an important component of the molecular machinery that harvests energy from glucose. Because this protein’s role in producing cellular energy is crucial, it has changed very little over millions of years. Protein sequencing has shown that there is a considerable amount of sequence similarity among cytochrome c molecules of different species; evolutionary relationships can be assessed by measuring the similarities or differences among various species’ protein sequences.
For example, scientists have determined that human cytochrome c contains 104 amino acids. For each cytochrome c molecule that has been sequenced to date from different organisms, 37 of these amino acids appear in the same position in each cytochrome c. This indicates that all of these organisms are descended from a common ancestor. On comparing the human and chimpanzee protein sequences, no sequence difference was found. When human and rhesus monkey sequences were compared, a single difference was found in one amino acid. In contrast, human-to-yeast comparisons show a difference in 44 amino acids, suggesting that humans

and chimpanzees have a more recent common ancestor than humans and the rhesus monkey, or humans and yeast.

Protein Structure

As discussed earlier, the shape of a protein is critical to its function. To understand how the protein gets its final shape or conformation, we need to understand the four levels of protein structure: primary, secondary, tertiary, and quaternary ([link]).

The unique sequence and number of amino acids in a polypeptide chain is its primary structure. The unique sequence for every protein is ultimately determined by the gene that encodes the protein. Any change in the gene sequence may lead to a different amino acid being added to the polypeptide chain, causing a change in protein structure and function. In sickle cell anemia, the hemoglobin β chain has a single amino acid substitution, causing a change in both the structure and function of the protein. What is most remarkable to consider is that a hemoglobin molecule is made up of two alpha chains and two beta chains that each consist of about 150 amino acids. The molecule, therefore, has about 600 amino acids. The structural difference between a normal hemoglobin molecule and a sickle cell molecule—that dramatically decreases life expectancy in the affected individuals—is a single amino acid of the 600.

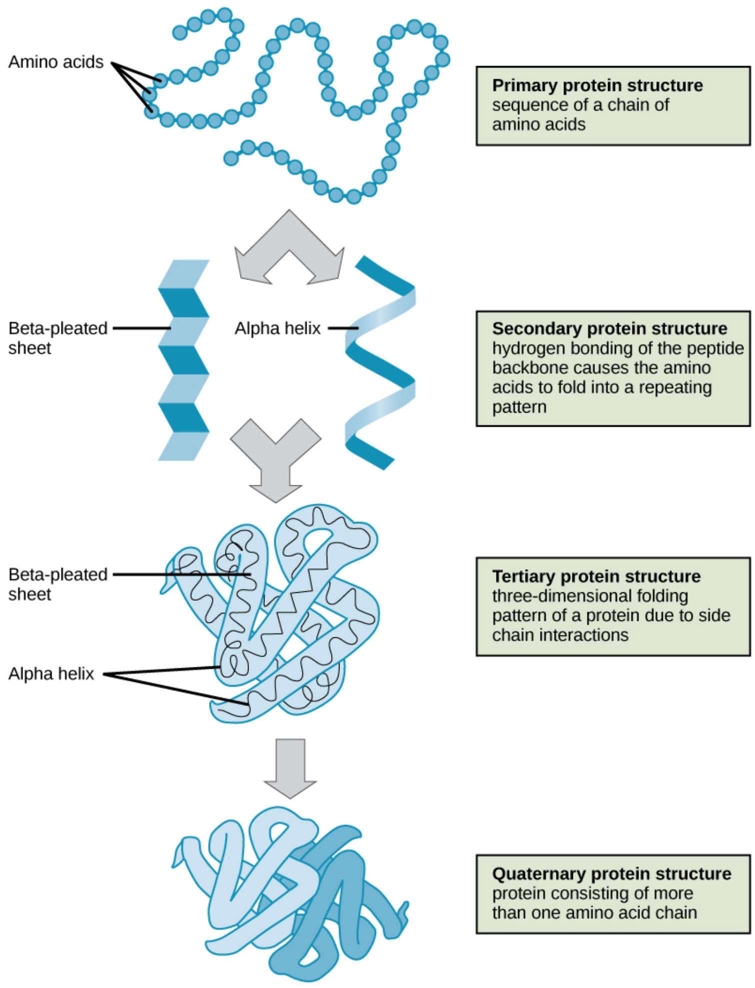
Because of this change of one amino acid in the chain, the normally biconcave, or disc-shaped, red blood cells assume a crescent or “sickle” shape, which clogs arteries. This can lead to a myriad of serious health problems, such as breathlessness, dizziness, headaches, and abdominal pain for those who have this disease.

Folding patterns resulting from interactions between the non-R group portions of amino acids give rise to the secondary structure of the protein. The most common are the alpha (α)-helix and beta (β)-pleated sheet structures. Both structures are held in shape by hydrogen bonds. In the alpha helix, the bonds form between every fourth amino acid and cause a twist in the amino acid chain.

In the β-pleated sheet, the “pleats” are formed by hydrogen bonding between atoms on the backbone of the polypeptide chain. The R groups are attached to the carbons, and extend above and below the folds of the pleat. The pleated segments align parallel to each other, and hydrogen bonds form between the same pairs of atoms on each of the aligned amino acids. The α-helix and β-pleated sheet structures are found in many globular and fibrous proteins.

The unique three-dimensional structure of a polypeptide is known as its tertiary structure. This structure is caused by chemical interactions between various amino acids and regions of the polypeptide. Primarily, the interactions among R groups create the complex three-dimensional tertiary structure of a protein. There may be ionic bonds formed between R groups on different amino acids, or hydrogen bonding beyond that involved in the secondary structure. When protein folding takes place, the hydrophobic R groups of nonpolar amino acids lay in the interior of the protein, whereas the hydrophilic R groups lay on the outside. The former types of interactions are also known as hydrophobic interactions.

In nature, some proteins are formed from several polypeptides, also known as subunits, and the interaction of these subunits forms the quaternary structure. Weak interactions between the subunits help to stabilize the overall structure. For example, hemoglobin is a combination of four polypeptide subunits.



The four levels of protein structure can be observed in these illustrations. (credit: modification of work by National Human Genome Research Institute)

Each protein has its own unique sequence and shape held together by chemical interactions. If the protein is subject to changes in temperature, pH, or exposure to chemicals, the protein structure may change, losing its shape in what is known as denaturation as discussed earlier. Denaturation is often reversible because the primary structure is preserved if the denaturing agent is removed, allowing the protein to resume its function. Sometimes denaturation is irreversible, leading to a loss of function. One example of

protein denaturation can be seen when an egg is fried or boiled. The albumin protein in the liquid egg white is denatured when placed in a hot pan, changing from a clear substance to an opaque white substance. Not all proteins are denatured at high temperatures; for instance, bacteria that survive in hot springs have proteins that are adapted to function at those temperatures.

Note:
Concept in Action

For an additional perspective on proteins, explore “Biomolecules: The Proteins” through this interactive animation.

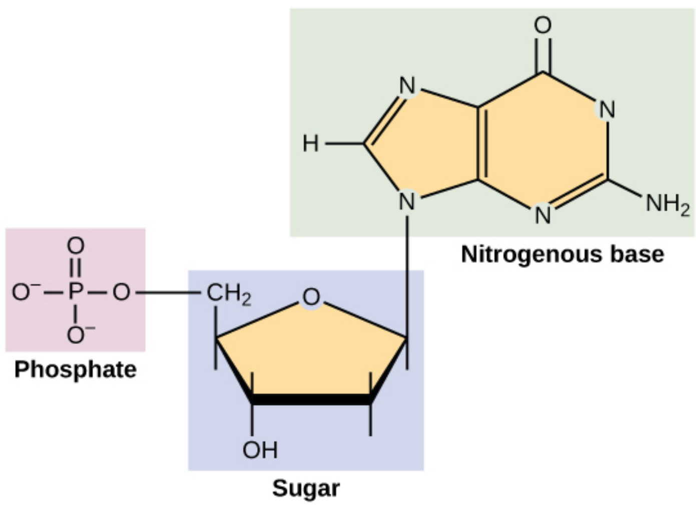
Nucleic Acids

Nucleic acids are key macromolecules in the continuity of life. They carry the genetic blueprint of a cell and carry instructions for the functioning of the cell.

The two main types of **nucleic acids** are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in all living organisms, ranging from single-celled bacteria to multicellular mammals.

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus, but instead use an RNA intermediary to communicate with the rest of the cell. Other types of RNA are also involved in protein synthesis and its regulation.

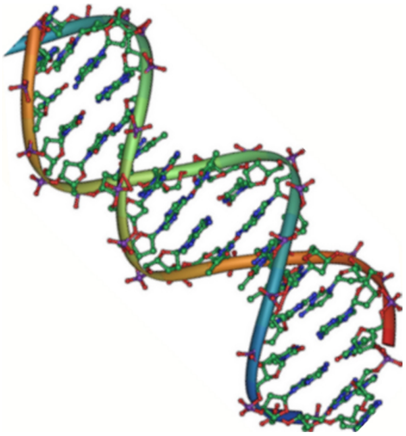
DNA and RNA are made up of monomers known as **nucleotides**. The nucleotides combine with each other to form a polynucleotide, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group ([link]). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to a phosphate group.



A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and a phosphate group.

DNA Double-Helical Structure

DNA has a double-helical structure ([link]). It is composed of two strands, or polymers, of nucleotides. The strands are formed with bonds between phosphate and sugar groups of adjacent nucleotides. The strands are bonded to each other at their bases with hydrogen bonds, and the strands coil about each other along their length, hence the "double helix" description, which means a double spiral.



The double-helix model shows DNA as two parallel strands of intertwining molecules. (credit: Jerome Walker, Dennis Myts)

The alternating sugar and phosphate groups lie on the outside of each strand, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, and these bases pair; the pairs are bound to each other by hydrogen bonds. The bases pair in such a way that the distance between the backbones of the two strands is the same all along the molecule.

Section Summary

Living things are carbon-based because carbon plays such a prominent role in the chemistry of living things. The four covalent bonding positions of the carbon atom can give rise to a wide diversity of compounds with many functions, accounting for the importance of carbon in living things. Carbohydrates are a group of macromolecules that are a vital energy source for the cell, provide structural support to many organisms, and can be found

on the surface of the cell as receptors or for cell recognition. Carbohydrates are classified as monosaccharides, disaccharides, and polysaccharides, depending on the number of monomers in the molecule.

Lipids are a class of macromolecules that are nonpolar and hydrophobic in nature. Major types include fats and oils, waxes, phospholipids, and steroids. Fats and oils are a stored form of energy and can include triglycerides. Fats and oils are usually made up of fatty acids and glycerol.

Proteins are a class of macromolecules that can perform a diverse range of functions for the cell. They help in metabolism by providing structural support and by acting as enzymes, carriers or as hormones. The building blocks of proteins are amino acids. Proteins are organized at four levels: primary, secondary, tertiary, and quaternary. Protein shape and function are intricately linked; any change in shape caused by changes in temperature, pH, or chemical exposure may lead to protein denaturation and a loss of function.

Nucleic acids are molecules made up of repeating units of nucleotides that direct cellular activities such as cell division and protein synthesis. Each nucleotide is made up of a pentose sugar, a nitrogenous base, and a phosphate group. There are two types of nucleic acids: DNA and RNA.

Multiple Choice

Exercise:

Problem: An example of a monosaccharide is ________.

a. fructose
b. glucose
c. galactose
d. all of the above

Solution:

D

Exercise:

Problem: Cellulose and starch are examples of _________.

a. monosaccharides
b. disaccharides
c. lipids
d. polysaccharides

Solution:

D

Exercise:

Problem: Phospholipids are important components of ___________.

a. the plasma membrane of cells
b. the ring structure of steroids
c. the waxy covering on leaves
d. the double bond in hydrocarbon chains

Solution:

A

Exercise:

Problem: The monomers that make up proteins are called __________.

a. nucleotides
b. disaccharides
c. amino acids
d. chaperones

Solution:

C

Free Response

Exercise:

Problem:

Explain at least three functions that lipids serve in plants and/or animals.

Solution:

Fat serves as a valuable way for animals to store energy. It can also provide insulation. Phospholipids and steroids are important components of cell membranes.

Exercise:

Problem:

Explain what happens if even one amino acid is substituted for another in a polypeptide chain. Provide a specific example.

Solution:

A change in gene sequence can lead to a different amino acid being added to a polypeptide chain instead of the normal one. This causes a change in protein structure and function. For example, in sickle cell anemia, the hemoglobin β chain has a single amino acid substitution. Because of this change, the disc-shaped red blood cells assume a crescent shape, which can result in serious health problems.

Glossary

amino acid
: a monomer of a protein

carbohydrate
: a biological macromolecule in which the ratio of carbon to hydrogen to oxygen is 1:2:1; carbohydrates serve as energy sources and structural support in cells

cellulose
: a polysaccharide that makes up the cell walls of plants and provides structural support to the cell

chitin
: a type of carbohydrate that forms the outer skeleton of arthropods, such as insects and crustaceans, and the cell walls of fungi

denaturation
: the loss of shape in a protein as a result of changes in temperature, pH, or exposure to chemicals

deoxyribonucleic acid (DNA)
: a double-stranded polymer of nucleotides that carries the hereditary information of the cell

disaccharide
: two sugar monomers that are linked together by a peptide bond

enzyme
: a catalyst in a biochemical reaction that is usually a complex or conjugated protein

fat
: a lipid molecule composed of three fatty acids and a glycerol (triglyceride) that typically exists in a solid form at room temperature

glycogen
: a storage carbohydrate in animals

hormone

a chemical signaling molecule, usually a protein or steroid, secreted by an endocrine gland or group of endocrine cells; acts to control or regulate specific physiological processes

lipids

a class of macromolecules that are nonpolar and insoluble in water

macromolecule

a large molecule, often formed by polymerization of smaller monomers

monosaccharide

a single unit or monomer of carbohydrates

nucleic acid

a biological macromolecule that carries the genetic information of a cell and carries instructions for the functioning of the cell

nucleotide

a monomer of nucleic acids; contains a pentose sugar, a phosphate group, and a nitrogenous base

oil

an unsaturated fat that is a liquid at room temperature

phospholipid

a major constituent of the membranes of cells; composed of two fatty acids and a phosphate group attached to the glycerol backbone

polypeptide

a long chain of amino acids linked by peptide bonds

polysaccharide

a long chain of monosaccharides; may be branched or unbranched

protein

a biological macromolecule composed of one or more chains of amino acids

ribonucleic acid (RNA)
a single-stranded polymer of nucleotides that is involved in protein synthesis

saturated fatty acid
a long-chain hydrocarbon with single covalent bonds in the carbon chain; the number of hydrogen atoms attached to the carbon skeleton is maximized

starch
a storage carbohydrate in plants

steroid
a type of lipid composed of four fused hydrocarbon rings

trans-fat
a form of unsaturated fat with the hydrogen atoms neighboring the double bond across from each other rather than on the same side of the double bond

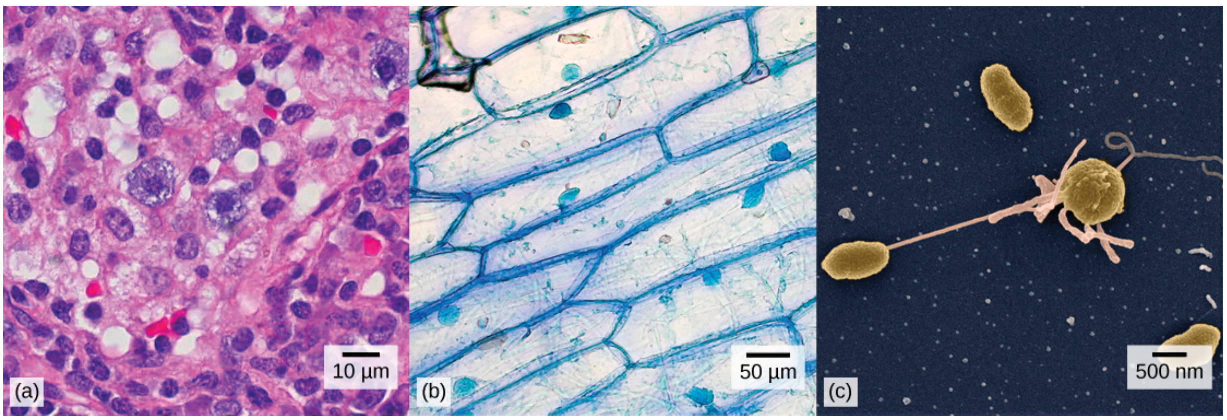
triglyceride
a fat molecule; consists of three fatty acids linked to a glycerol molecule

unsaturated fatty acid
a long-chain hydrocarbon that has one or more than one double bonds in the hydrocarbon chain

Introduction class="introduction"

(a) Nasal sinus cells (viewed with a light microscope), (b) onion cells (viewed with a light microscope), and (c) *Vibrio tasmaniensis* bacterial cells (viewed using a scanning electron microscope) are from very different organisms, yet all share certain characteristics of basic cell structure. (credit a: modification of work by Ed Uthman, MD; credit b: modification of work by Umberto Salvagnin; credit c:

modification of work by Anthony D'Onofrio; scale-bar data from Matt Russell)



Close your eyes and picture a brick wall. What is the basic building block of that wall? It is a single brick, of course. Like a brick wall, your body is composed of basic building blocks, and the building blocks of your body are cells.

Your body has many kinds of cells, each specialized for a specific purpose. Just as a home is made from a variety of building materials, the human body is constructed from many cell types. For example, epithelial cells protect the surface of the body and cover the organs and body cavities within. Bone cells help to support and protect the body. Cells of the immune system fight invading bacteria. Additionally, red blood cells carry oxygen throughout the body. Each of these cell types plays a vital role during the growth, development, and day-to-day maintenance of the body. In spite of their enormous variety, however, all cells share certain fundamental characteristics.

How Cells Are Studied

By the end of this section, you will be able to:

- Describe the roles of cells in organisms
- Compare and contrast light microscopy and electron microscopy
- Summarize the cell theory

A cell is the smallest unit of a living thing. A living thing, like you, is called an organism. Thus, cells are the basic building blocks of all organisms.

In multicellular organisms, several cells of one particular kind interconnect with each other and perform shared functions to form tissues (for example, muscle tissue, connective tissue, and nervous tissue), several tissues combine to form an organ (for example, stomach, heart, or brain), and several organs make up an organ system (such as the digestive system, circulatory system, or nervous system). Several systems functioning together form an organism (such as an elephant, for example).

There are many types of cells, and all are grouped into one of two broad categories: prokaryotic and eukaryotic. Animal cells, plant cells, fungal cells, and protist cells are classified as eukaryotic, whereas bacteria and archaea cells are classified as prokaryotic. Before discussing the criteria for determining whether a cell is prokaryotic or eukaryotic, let us first examine how biologists study cells.

Microscopy

Cells vary in size. With few exceptions, individual cells are too small to be seen with the naked eye, so scientists use microscopes to study them. A **microscope** is an instrument that magnifies an object. Most images of cells are taken with a microscope and are called micrographs.

Light Microscopes

To give you a sense of the size of a cell, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as µm)

in diameter; the head of a pin is about two thousandths of a meter (millimeters, or mm) in diameter. That means that approximately 250 red blood cells could fit on the head of a pin.

The optics of the lenses of a light microscope changes the orientation of the image. A specimen that is right-side up and facing right on the microscope slide will appear upside-down and facing left when viewed through a microscope, and vice versa. Similarly, if the slide is moved left while looking through the microscope, it will appear to move right, and if moved down, it will seem to move up. This occurs because microscopes use two sets of lenses to magnify the image. Due to the manner in which light travels through the lenses, this system of lenses produces an inverted image (binoculars and a dissecting microscope work in a similar manner, but include an additional magnification system that makes the final image appear to be upright).

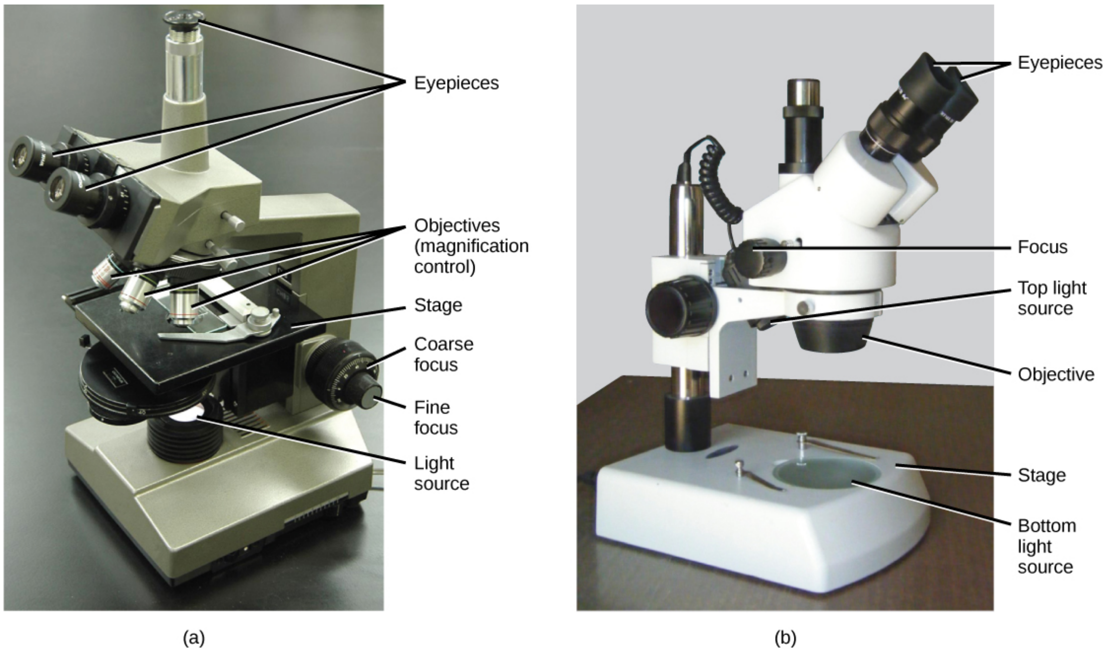
Most student microscopes are classified as light microscopes ([link]**a**). Visible light both passes through and is bent by the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains. Staining, however, usually kills the cells.

Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the degree of enlargement of an object. Resolving power is the ability of a microscope to allow the eye to distinguish two adjacent structures as separate; the higher the resolution, the closer those two objects can be, and the better the clarity and detail of the image. When oil immersion lenses are used, magnification is usually increased to 1,000 times for the study of smaller cells, like most prokaryotic cells. Because light entering a specimen from below is focused onto the eye of an observer, the specimen can be viewed using light microscopy. For this reason, for light to pass through a specimen, the sample must be thin or translucent.

Note:
Concept in Action

For another perspective on cell size, try the HowBig interactive.

A second type of microscope used in laboratories is the dissecting microscope ([link]**b**). These microscopes have a lower magnification (20 to 80 times the object size) than light microscopes and can provide a three-dimensional view of the specimen. Thick objects can be examined with many components in focus at the same time. These microscopes are designed to give a magnified and clear view of tissue structure as well as the anatomy of the whole organism. Like light microscopes, most modern dissecting microscopes are also binocular, meaning that they have two separate lens systems, one for each eye. The lens systems are separated by a certain distance, and therefore provide a sense of depth in the view of their subject to make manipulations by hand easier. Dissecting microscopes also have optics that correct the image so that it appears as if being seen by the naked eye and not as an inverted image. The light illuminating a sample under a dissecting microscope typically comes from above the sample, but may also be directed from below.



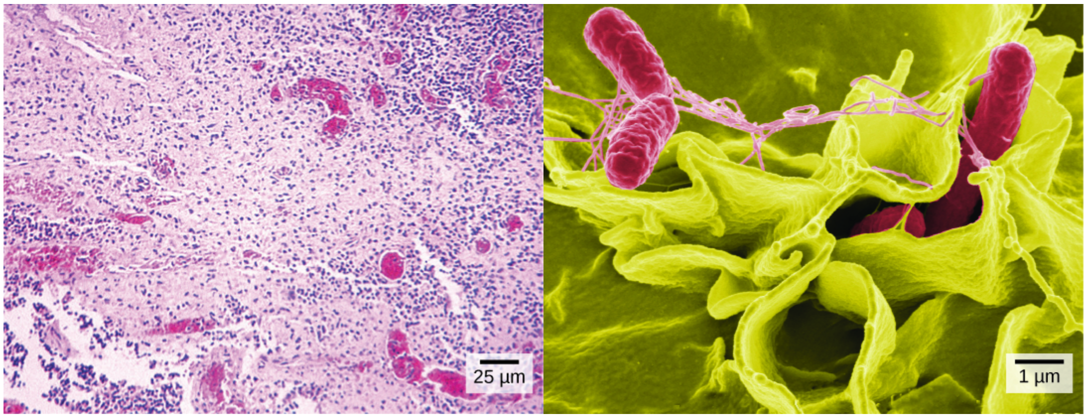
(a) Most light microscopes used in a college biology lab can magnify cells up to approximately 400 times. (b) Dissecting microscopes have a lower magnification than light microscopes and are used to examine larger objects, such as tissues.

Electron Microscopes

In contrast to light microscopes, electron microscopes use a beam of electrons instead of a beam of light. Not only does this allow for higher magnification and, thus, more detail ([link]), it also provides higher resolving power. Preparation of a specimen for viewing under an electron microscope will kill it; therefore, live cells cannot be viewed using this type of microscopy. In addition, the electron beam moves best in a vacuum, making it impossible to view living materials.

In a scanning electron microscope, a beam of electrons moves back and forth across a cell's surface, rendering the details of cell surface characteristics by reflection. Cells and other structures are usually coated

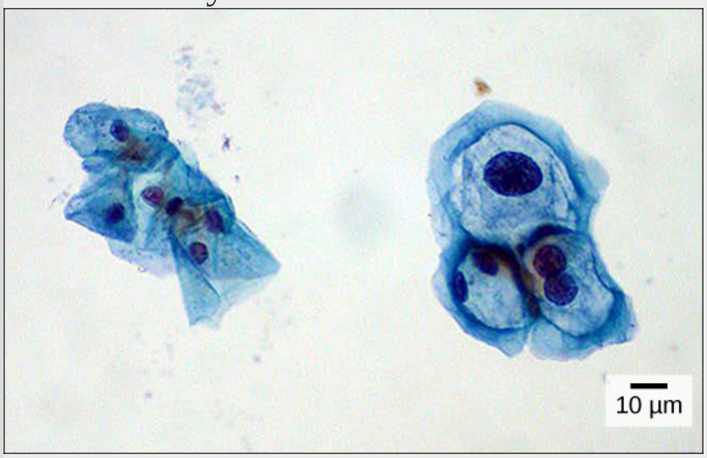
with a metal like gold. In a transmission electron microscope, the electron beam is transmitted through the cell and provides details of a cell's internal structures. As you might imagine, electron microscopes are significantly more bulky and expensive than are light microscopes.



(a) *Salmonella* bacteria are viewed with a light microscope. (b) This scanning electron micrograph shows *Salmonella* bacteria (in red) invading human cells. (credit a: modification of work by CDC, Armed Forces Institute of Pathology, Charles N. Farmer; credit b: modification of work by Rocky Mountain Laboratories, NIAID, NIH; scale-bar data from Matt Russell)

Note:
Careers in Action
Cytotechnologist
Have you ever heard of a medical test called a Pap smear ([link])? In this test, a doctor takes a small sample of cells from the uterine cervix of a patient and sends it to a medical lab where a cytotechnologist stains the cells and examines them for any changes that could indicate cervical cancer or a microbial infection.

Cytotechnologists (*cyto*- = cell) are professionals who study cells through microscopic examinations and other laboratory tests. They are trained to determine which cellular changes are within normal limits or are abnormal. Their focus is not limited to cervical cells; they study cellular specimens that come from all organs. When they notice abnormalities, they consult a pathologist, who is a medical doctor who can make a clinical diagnosis. Cytotechnologists play vital roles in saving people's lives. When abnormalities are discovered early, a patient's treatment can begin sooner, which usually increases the chances of successful treatment.



These uterine cervix cells, viewed through a light microscope, were obtained from a Pap smear. Normal cells are on the left. The cells on the right are infected with human papillomavirus. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

Cell Theory

The microscopes we use today are far more complex than those used in the 1600s by Antony van Leeuwenhoek, a Dutch shopkeeper who had great

skill in crafting lenses. Despite the limitations of his now-ancient lenses, van Leeuwenhoek observed the movements of protists (a type of single-celled organism) and sperm, which he collectively termed “animalcules.”

In a 1665 publication called *Micrographia*, experimental scientist Robert Hooke coined the term “cell” (from the Latin *cella*, meaning “small room”) for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses and microscope construction enabled other scientists to see different components inside cells.

By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the **unified cell theory**, which states that all living things are composed of one or more cells, that the cell is the basic unit of life, and that all new cells arise from existing cells. These principles still stand today.

Section Summary

A cell is the smallest unit of life. Most cells are so small that they cannot be viewed with the naked eye. Therefore, scientists must use microscopes to study cells. Electron microscopes provide higher magnification, higher resolution, and more detail than light microscopes. The unified cell theory states that all organisms are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells.

Multiple Choice

Exercise:

Problem:

When viewing a specimen through a light microscope, scientists use __________ to distinguish the individual components of cells.

a. a beam of electrons
b. radioactive isotopes

c. special stains
d. high temperatures

Solution:

C

Exercise:

Problem: The ____________ is the basic unit of life.

a. organism
b. cell
c. tissue
d. organ

Solution:

B

Free Response

Exercise:

Problem:

What are the advantages and disadvantages of light, transmission, and scanning electron microscopes?

Solution:

The advantages of light microscopes are that they are easily obtained, and the light beam does not kill the cells. However, typical light microscopes are somewhat limited in the amount of detail that they can reveal. Electron microscopes are ideal because you can view intricate details, but they are bulky and costly, and preparation for the

microscopic examination kills the specimen. Transmission electron microscopes are designed to examine the internal structures of a cell, whereas a scanning electron microscope only allows visualization of the surface of a structure.

Glossary

microscope
: the instrument that magnifies an object

unified cell theory
: the biological concept that states that all organisms are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells

Comparing Prokaryotic and Eukaryotic Cells
By the end of this section, you will be able to:

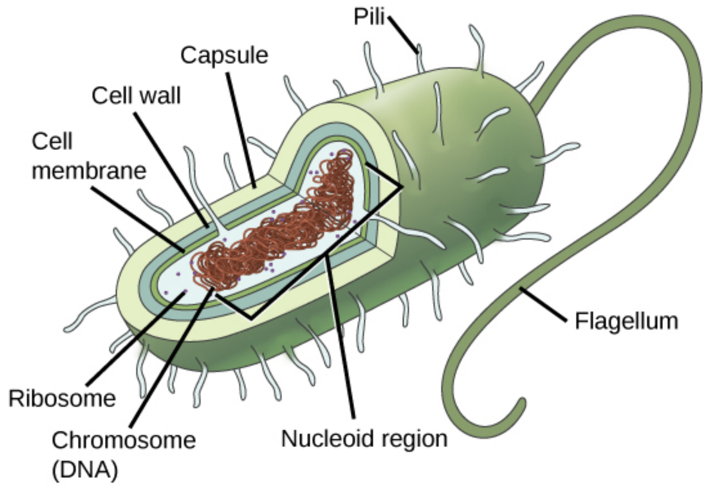
- Name examples of prokaryotic and eukaryotic organisms
- Compare and contrast prokaryotic cells and eukaryotic cells
- Describe the relative sizes of different kinds of cells

Cells fall into one of two broad categories: prokaryotic and eukaryotic. The predominantly single-celled organisms of the domains Bacteria and Archaea are classified as prokaryotes (*pro*- = before; -*karyon*- = nucleus). Animal cells, plant cells, fungi, and protists are eukaryotes (*eu*- = true).

Components of Prokaryotic Cells

All cells share four common components: 1) a plasma membrane, an outer covering that separates the cell's interior from its surrounding environment; 2) cytoplasm, consisting of a jelly-like region within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) ribosomes, particles that synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A **prokaryotic cell** is a simple, single-celled (unicellular) organism that lacks a nucleus, or any other membrane-bound organelle. We will shortly come to see that this is significantly different in eukaryotes. Prokaryotic DNA is found in the central part of the cell: a darkened region called the nucleoid ([link]).



This figure shows the generalized structure of a prokaryotic cell.

Unlike Archaea and eukaryotes, bacteria have a cell wall made of peptidoglycan, comprised of sugars and amino acids, and many have a polysaccharide capsule ([link]). The cell wall acts as an extra layer of protection, helps the cell maintain its shape, and prevents dehydration. The capsule enables the cell to attach to surfaces in its environment. Some prokaryotes have flagella, pili, or fimbriae. Flagella are used for locomotion, while most pili are used to exchange genetic material during a type of reproduction called conjugation.

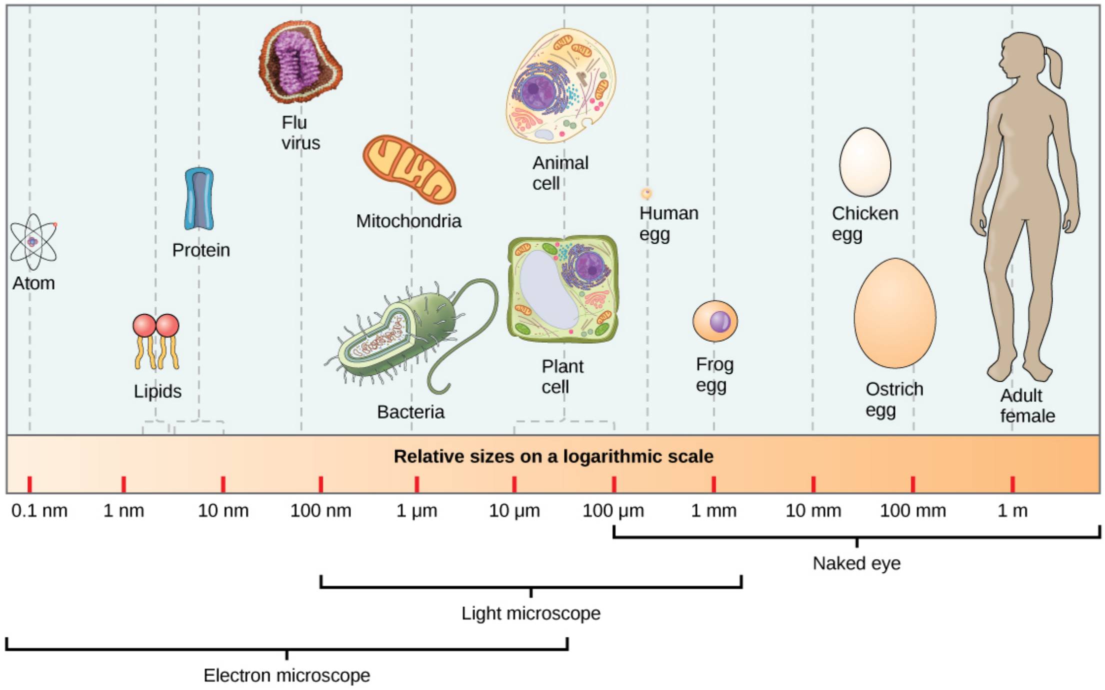
Eukaryotic Cells

In nature, the relationship between form and function is apparent at all levels, including the level of the cell, and this will become clear as we explore eukaryotic cells. The principle “form follows function” is found in many contexts. For example, birds and fish have streamlined bodies that allow them to move quickly through the medium in which they live, be it air or water. It means that, in general, one can deduce the function of a structure by looking at its form, because the two are matched.

A **eukaryotic cell** is a cell that has a membrane-bound nucleus and other membrane-bound compartments or sacs, called **organelles**, which have specialized functions. The word eukaryotic means “true kernel” or “true nucleus,” alluding to the presence of the membrane-bound nucleus in these cells. The word “organelle” means “little organ,” and, as already mentioned, organelles have specialized cellular functions, just as the organs of your body have specialized functions.

Cell Size

At 0.1–5.0 µm in diameter, prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10–100 µm ([link]). The small size of prokaryotes allows ions and organic molecules that enter them to quickly spread to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly move out. However, larger eukaryotic cells have evolved different structural adaptations to enhance cellular transport. Indeed, the large size of these cells would not be possible without these adaptations. In general, cell size is limited because volume increases much more quickly than does cell surface area. As a cell becomes larger, it becomes more and more difficult for the cell to acquire sufficient materials to support the processes inside the cell, because the relative size of the surface area across which materials must be transported declines.



This figure shows the relative sizes of different kinds of cells and cellular components. An adult human is shown for comparison.

Section Summary

Prokaryotes are predominantly single-celled organisms of the domains Bacteria and Archaea. All prokaryotes have plasma membranes, cytoplasm, ribosomes, a cell wall, DNA, and lack membrane-bound organelles. Many also have polysaccharide capsules. Prokaryotic cells range in diameter from 0.1–5.0 µm.

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. Eukaryotic cells tend to be 10 to 100 times the size of prokaryotic cells.

Multiple Choice

Exercise:

Problem: Which of these do all prokaryotes and eukaryotes share?

a. nuclear envelope
b. cell walls
c. organelles
d. plasma membrane

Solution:

D

Exercise:

Problem:

A typical prokaryotic cell ___________________ compared to a eukaryotic cell.

a. is smaller in size by a factor of 100
b. is similar in size
c. is smaller in size by a factor of one million
d. is larger in size by a factor of 10

Solution:

A

Free Response

Exercise:

Problem:

Describe the structures that are characteristic of a prokaryote cell.

Solution:

Prokaryotic cells are surrounded by a plasma membrane and have DNA, cytoplasm, and ribosomes, like eukaryotic cells. They also have cell walls and may have a cell capsule. Prokaryotes have a single large chromosome that is not surrounded by a nuclear membrane. Prokaryotes may have flagella or motility, pili for conjugation, and fimbriae for adhesion to surfaces.

Glossary

eukaryotic cell
: a cell that has a membrane-bound nucleus and several other membrane-bound compartments or sacs

organelle
: a membrane-bound compartment or sac within a cell

prokaryotic cell
: a unicellular organism that lacks a nucleus or any other membrane-bound organelle

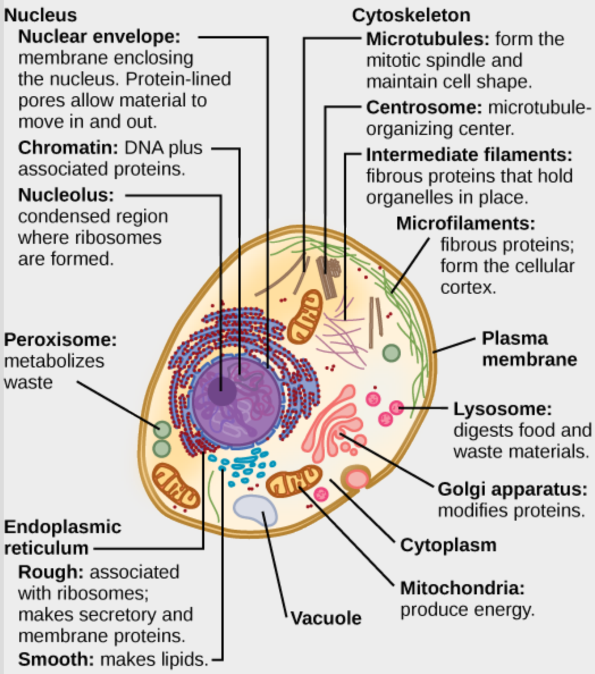
Eukaryotic Cells
By the end of this section, you will be able to:

- Describe the structure of eukaryotic plant and animal cells
- State the role of the plasma membrane
- Summarize the functions of the major cell organelles
- Describe the cytoskeleton and extracellular matrix

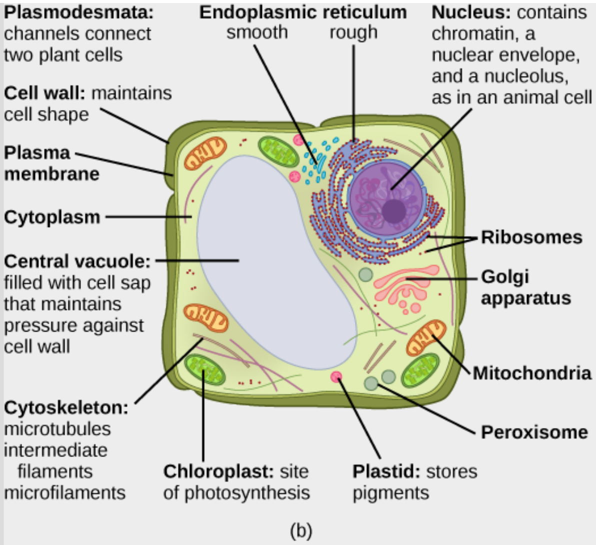
At this point, it should be clear that eukaryotic cells have a more complex structure than do prokaryotic cells. Organelles allow for various functions to occur in the cell at the same time. Before discussing the functions of organelles within a eukaryotic cell, let us first examine two important components of the cell: the plasma membrane and the cytoplasm.

Note:
Art Connection

This figure shows (a) a typical animal cell and (b) a typical plant cell.



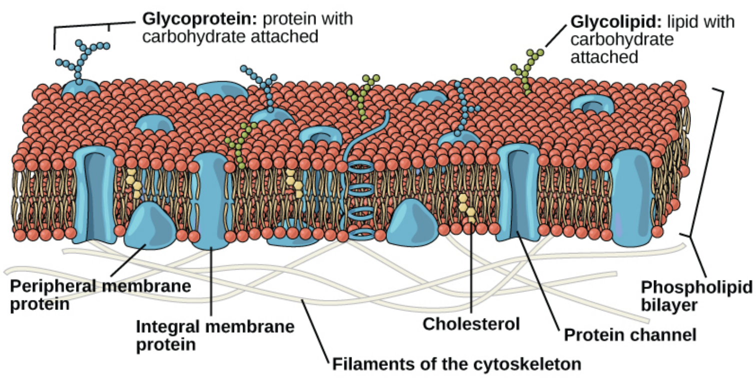
(a)



What structures does a plant cell have that an animal cell does not have? What structures does an animal cell have that a plant cell does not have?

The Plasma Membrane

Like prokaryotes, eukaryotic cells have a **plasma membrane** ([link]) made up of a phospholipid bilayer with embedded proteins that separates the internal contents of the cell from its surrounding environment. A phospholipid is a lipid molecule composed of two fatty acid chains, a glycerol backbone, and a phosphate group. The plasma membrane regulates the passage of some substances, such as organic molecules, ions, and water, preventing the passage of some to maintain internal conditions, while actively bringing in or removing others. Other compounds move passively across the membrane.



The plasma membrane is a phospholipid bilayer with embedded proteins. There are other components, such as cholesterol and

carbohydrates, which can be found in the membrane in addition to phospholipids and protein.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (singular = microvillus). This folding increases the surface area of the plasma membrane. Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. This is an excellent example of form matching the function of a structure.

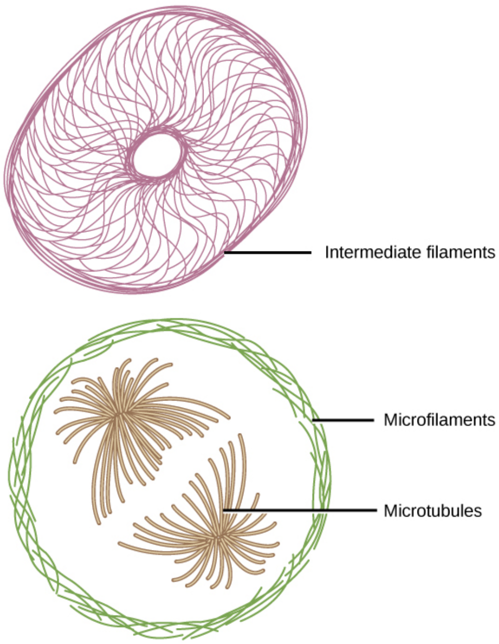
People with celiac disease have an immune response to gluten, which is a protein found in wheat, barley, and rye. The immune response damages microvilli, and thus, afflicted individuals cannot absorb nutrients. This leads to malnutrition, cramping, and diarrhea. Patients suffering from celiac disease must follow a gluten-free diet.

The Cytoplasm

The **cytoplasm** comprises the contents of a cell between the plasma membrane and the nuclear envelope (a structure to be discussed shortly). It is made up of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals ([link]). Even though the cytoplasm consists of 70 to 80 percent water, it has a semi-solid consistency, which comes from the proteins within it. However, proteins are not the only organic molecules found in the cytoplasm. Glucose and other simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, and derivatives of glycerol are found there too. Ions of sodium, potassium, calcium, and many other elements are also dissolved in the cytoplasm. Many metabolic reactions, including protein synthesis, take place in the cytoplasm.

The Cytoskeleton

If you were to remove all the organelles from a cell, would the plasma membrane and the cytoplasm be the only components left? No. Within the cytoplasm, there would still be ions and organic molecules, plus a network of protein fibers that helps to maintain the shape of the cell, secures certain organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move independently. Collectively, this network of protein fibers is known as the **cytoskeleton**. There are three types of fibers within the cytoskeleton: microfilaments, also known as actin filaments, intermediate filaments, and microtubules ([link]).



Microfilaments, intermediate filaments, and microtubules compose a cell’s cytoskeleton.

Microfilaments are the thinnest of the cytoskeletal fibers and function in moving cellular components, for example, during cell division. They also maintain the structure of microvilli, the extensive folding of the plasma membrane found in cells dedicated to absorption. These components are also common in muscle cells and are responsible for muscle cell contraction. Intermediate filaments are of intermediate diameter and have structural functions, such as maintaining the shape of the cell and anchoring organelles. Keratin, the compound that strengthens hair and nails, forms one type of intermediate filament. Microtubules are the thickest of the cytoskeletal fibers. These are hollow tubes that can dissolve and reform quickly. Microtubules guide organelle movement and are the structures that pull chromosomes to their poles during cell division. They are also the structural components of flagella and cilia. In cilia and flagella, the microtubules are organized as a circle of nine double microtubules on the outside and two microtubules in the center.

The centrosome is a region near the nucleus of animal cells that functions as a microtubule-organizing center. It contains a pair of centrioles, two structures that lie perpendicular to each other. Each centriole is a cylinder of nine triplets of microtubules.

The centrosome replicates itself before a cell divides, and the centrioles play a role in pulling the duplicated chromosomes to opposite ends of the dividing cell. However, the exact function of the centrioles in cell division is not clear, since cells that have the

centrioles removed can still divide, and plant cells, which lack centrioles, are capable of cell division.

Flagella and Cilia

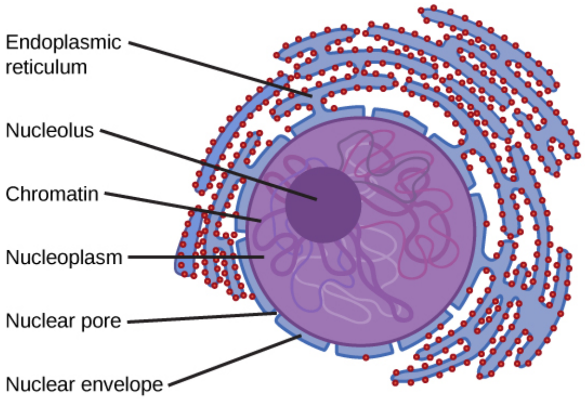
Flagella (singular = flagellum) are long, hair-like structures that extend from the plasma membrane and are used to move an entire cell, (for example, sperm, *Euglena*). When present, the cell has just one flagellum or a few flagella. When **cilia** (singular = cilium) are present, however, they are many in number and extend along the entire surface of the plasma membrane. They are short, hair-like structures that are used to move entire cells (such as paramecium) or move substances along the outer surface of the cell (for example, the cilia of cells lining the fallopian tubes that move the ovum toward the uterus, or cilia lining the cells of the respiratory tract that move particulate matter toward the throat that mucus has trapped).

The Endomembrane System

The **endomembrane system** (*endo* = within) is a group of membranes and organelles ([link]) in eukaryotic cells that work together to modify, package, and transport lipids and proteins. It includes the nuclear envelope, lysosomes, and vesicles, the endoplasmic reticulum and Golgi apparatus, which we will cover shortly. Although not technically *within* the cell, the plasma membrane is included in the endomembrane system because, as you will see, it interacts with the other endomembranous organelles.

The Nucleus

Typically, the nucleus is the most prominent organelle in a cell ([link]). The **nucleus** (plural = nuclei) houses the cell's DNA in the form of chromatin and directs the synthesis of ribosomes and proteins. Let us look at it in more detail ([link]).



The outermost boundary of the nucleus is the nuclear envelope. Notice that the nuclear envelope consists of two phospholipid bilayers (membranes)—an outer membrane and an inner membrane—in contrast to the plasma membrane ([link]), which consists of only one phospholipid bilayer. (credit: modification of work by NIGMS, NIH)

The **nuclear envelope** is a double-membrane structure that constitutes the outermost portion of the nucleus ([link]). Both the inner and outer membranes of the nuclear envelope are phospholipid bilayers.

The nuclear envelope is punctuated with pores that control the passage of ions, molecules, and RNA between the nucleoplasm and the cytoplasm.

To understand chromatin, it is helpful to first consider chromosomes. Chromosomes are structures within the nucleus that are made up of DNA, the hereditary material, and proteins. This combination of DNA and proteins is called chromatin. In eukaryotes, chromosomes are linear structures. Every species has a specific number of chromosomes in the nucleus of its body cells. For example, in humans, the chromosome number is 46, whereas in fruit flies, the chromosome number is eight.

Chromosomes are only visible and distinguishable from one another when the cell is getting ready to divide. When the cell is in the growth and maintenance phases of its life cycle, the chromosomes resemble an unwound, jumbled bunch of threads.

We already know that the nucleus directs the synthesis of ribosomes, but how does it do this? Some chromosomes have sections of DNA that encode ribosomal RNA. A darkly staining area within the nucleus, called the **nucleolus** (plural = nucleoli), aggregates the ribosomal RNA with associated proteins to assemble the ribosomal subunits that are then transported through the nuclear pores into the cytoplasm.

The Endoplasmic Reticulum

The **endoplasmic reticulum (ER)** ([link]) is a series of interconnected membranous tubules that collectively modify proteins and synthesize lipids. However, these two functions are performed in separate areas of the endoplasmic reticulum: the rough endoplasmic reticulum and the smooth endoplasmic reticulum, respectively.

The hollow portion of the ER tubules is called the lumen or cisternal space. The membrane of the ER, which is a phospholipid bilayer embedded with proteins, is continuous with the nuclear envelope.

The **rough endoplasmic reticulum (RER)** is so named because the ribosomes attached to its cytoplasmic surface give it a studded appearance when viewed through an electron microscope.

The ribosomes synthesize proteins while attached to the ER, resulting in transfer of their newly synthesized proteins into the lumen of the RER where they undergo modifications such as folding or addition of sugars. The RER also makes phospholipids for cell membranes.

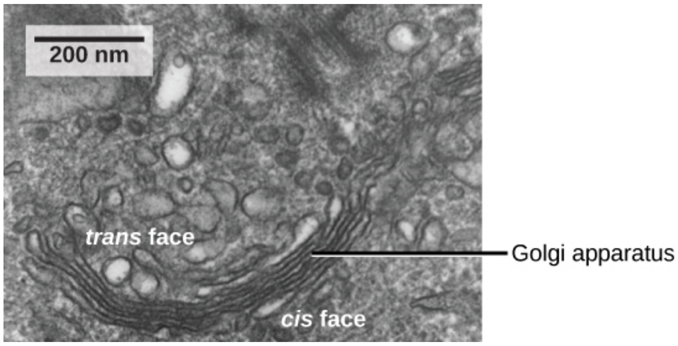
If the phospholipids or modified proteins are not destined to stay in the RER, they will be packaged within vesicles and transported from the RER by budding from the membrane ([link]). Since the RER is engaged in modifying proteins that will be secreted from the cell, it is abundant in cells that secrete proteins, such as the liver.

The **smooth endoplasmic reticulum (SER)** is continuous with the RER but has few or no ribosomes on its cytoplasmic surface (see [link]). The SER's functions include synthesis of carbohydrates, lipids (including phospholipids), and steroid hormones; detoxification of medications and poisons; alcohol metabolism; and storage of calcium ions.

The Golgi Apparatus

We have already mentioned that vesicles can bud from the ER, but where do the vesicles go? Before reaching their final destination, the lipids or proteins within the transport vesicles need to be sorted, packaged, and tagged so that they wind up in the right place. The sorting, tagging, packaging, and distribution of lipids and proteins take place in the

Golgi apparatus (also called the Golgi body), a series of flattened membranous sacs ([link]).



The Golgi apparatus in this transmission electron micrograph of a white blood cell is visible as a stack of semicircular flattened rings in the lower portion of this image. Several vesicles can be seen near the Golgi apparatus. (credit: modification of work by Louisa Howard; scale-bar data from Matt Russell)

The Golgi apparatus has a receiving face near the endoplasmic reticulum and a releasing face on the side away from the ER, toward the cell membrane. The transport vesicles that form from the ER travel to the receiving face, fuse with it, and empty their contents into the lumen of the Golgi apparatus. As the proteins and lipids travel through the Golgi, they undergo further modifications. The most frequent modification is the addition of short chains of sugar molecules. The newly modified proteins and lipids are then tagged with small molecular groups to enable them to be routed to their proper destinations.

Finally, the modified and tagged proteins are packaged into vesicles that bud from the opposite face of the Golgi. While some of these vesicles, transport vesicles, deposit their contents into other parts of the cell where they will be used, others, secretory vesicles, fuse with the plasma membrane and release their contents outside the cell.

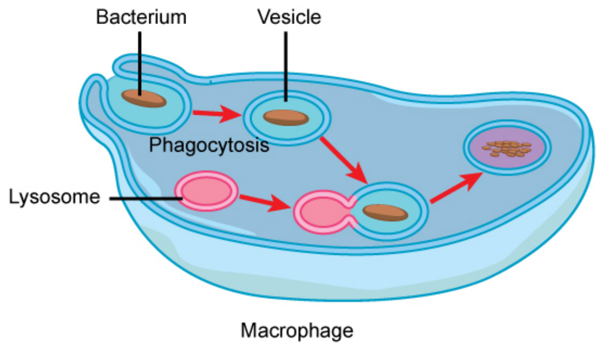
The amount of Golgi in different cell types again illustrates that form follows function within cells. Cells that engage in a great deal of secretory activity (such as cells of the salivary glands that secrete digestive enzymes or cells of the immune system that secrete antibodies) have an abundant number of Golgi.

In plant cells, the Golgi has an additional role of synthesizing polysaccharides, some of which are incorporated into the cell wall and some of which are used in other parts of the cell.

Lysosomes

In animal cells, the **lysosomes** are the cell's "garbage disposal." Digestive enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles. In single-celled eukaryotes, lysosomes are important for digestion of the food they ingest and the recycling of organelles. These enzymes are active at a much lower pH (more acidic) than those located in the cytoplasm. Many reactions that take place in the cytoplasm could not occur at a low pH, thus the advantage of compartmentalizing the eukaryotic cell into organelles is apparent.

Lysosomes also use their hydrolytic enzymes to destroy disease-causing organisms that might enter the cell. A good example of this occurs in a group of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis, a section of the plasma membrane of the macrophage invaginates (folds in) and engulfs a pathogen. The invaginated section, with the pathogen inside, then pinches itself off from the plasma membrane and becomes a vesicle. The vesicle fuses with a lysosome. The lysosome's hydrolytic enzymes then destroy the pathogen ([link]).

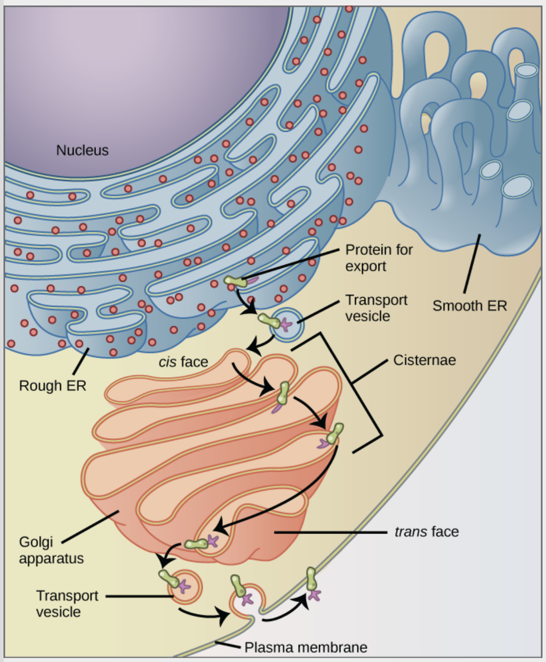


A macrophage has phagocytized a potentially pathogenic bacterium into a vesicle, which then fuses with a lysosome within the cell so that the pathogen can be destroyed. Other organelles are present in the cell, but for simplicity, are not shown.

Vesicles and Vacuoles

Vesicles and **vacuoles** are membrane-bound sacs that function in storage and transport. Vacuoles are somewhat larger than vesicles, and the membrane of a vacuole does not fuse with the membranes of other cellular components. Vesicles can fuse with other membranes within the cell system. Additionally, enzymes within plant vacuoles can break down macromolecules.

Note:
Art Connection



The endomembrane system works to modify, package, and transport lipids and proteins. (credit: modification of work by Magnus Manske)

Why does the *cis* face of the Golgi not face the plasma membrane?

Ribosomes

Ribosomes are the cellular structures responsible for protein synthesis. When viewed through an electron microscope, free ribosomes appear as either clusters or single tiny dots

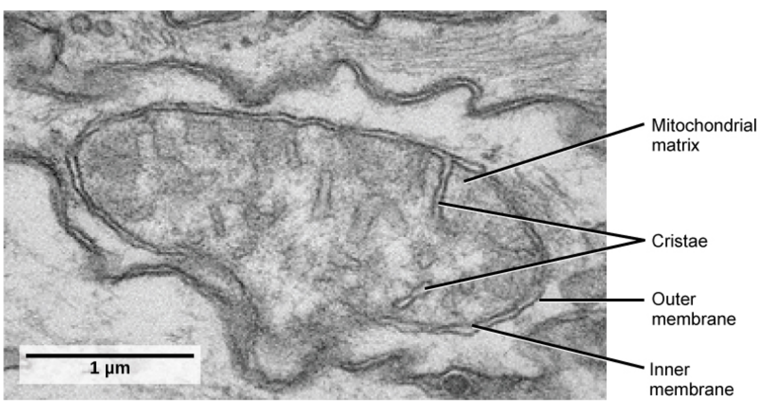
floating freely in the cytoplasm. Ribosomes may be attached to either the cytoplasmic side of the plasma membrane or the cytoplasmic side of the endoplasmic reticulum ([link]). Electron microscopy has shown that ribosomes consist of large and small subunits. Ribosomes are enzyme complexes that are responsible for protein synthesis.

Because protein synthesis is essential for all cells, ribosomes are found in practically every cell, although they are smaller in prokaryotic cells. They are particularly abundant in immature red blood cells for the synthesis of hemoglobin, which functions in the transport of oxygen throughout the body.

Mitochondria

Mitochondria (singular = mitochondrion) are often called the “powerhouses” or “energy factories” of a cell because they are responsible for making adenosine triphosphate (ATP), the cell’s main energy-carrying molecule. The formation of ATP from the breakdown of glucose is known as cellular respiration. Mitochondria are oval-shaped, double-membrane organelles ([link]) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The inner layer has folds called cristae, which increase the surface area of the inner membrane. The area surrounded by the folds is called the mitochondrial matrix. The cristae and the matrix have different roles in cellular respiration.

In keeping with our theme of form following function, it is important to point out that muscle cells have a very high concentration of mitochondria because muscle cells need a lot of energy to contract.



This transmission electron micrograph shows a mitochondrion as viewed with an electron microscope. Notice the inner and outer membranes, the cristae, and the mitochondrial matrix. (credit: modification of work by Matthew Britton; scale-bar data from Matt Russell)

Peroxisomes

Peroxisomes are small, round organelles enclosed by single membranes. They carry out oxidation reactions that break down fatty acids and amino acids. They also detoxify many poisons that may enter the body. Alcohol is detoxified by peroxisomes in liver cells. A byproduct of these oxidation reactions is hydrogen peroxide, H_2O_2, which is contained within the peroxisomes to prevent the chemical from causing damage to cellular components outside of the organelle. Hydrogen peroxide is safely broken down by peroxisomal enzymes into water and oxygen.

Animal Cells versus Plant Cells

Despite their fundamental similarities, there are some striking differences between animal and plant cells (see [link]). Animal cells have centrioles, centrosomes (discussed under the cytoskeleton), and lysosomes, whereas plant cells do not. Plant cells have a cell wall, chloroplasts, plasmodesmata, and plastids used for storage, and a large central vacuole, whereas animal cells do not.

The Cell Wall

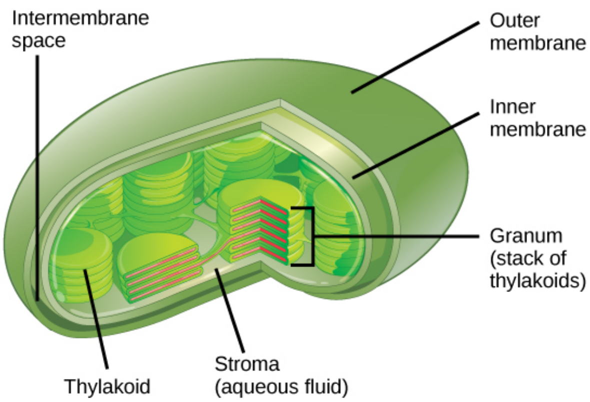
In [link]**b**, the diagram of a plant cell, you see a structure external to the plasma membrane called the cell wall. The **cell wall** is a rigid covering that protects the cell, provides structural support, and gives shape to the cell. Fungal and protist cells also have cell walls.

While the chief component of prokaryotic cell walls is peptidoglycan, the major organic molecule in the plant cell wall is cellulose, a polysaccharide made up of long, straight chains of glucose units. When nutritional information refers to dietary fiber, it is referring to the cellulose content of food.

Chloroplasts

Like mitochondria, chloroplasts also have their own DNA and ribosomes. **Chloroplasts** function in photosynthesis and can be found in eukaryotic cells such as plants and algae. In photosynthesis, carbon dioxide, water, and light energy are used to make glucose and oxygen. This is the major difference between plants and animals: Plants (autotrophs) are able to make their own food, like glucose, whereas animals (heterotrophs) must rely on other organisms for their organic compounds or food source.

Like mitochondria, chloroplasts have outer and inner membranes, but within the space enclosed by a chloroplast's inner membrane is a set of interconnected and stacked, fluid-filled membrane sacs called thylakoids ([link]). Each stack of thylakoids is called a granum (plural = grana). The fluid enclosed by the inner membrane and surrounding the grana is called the stroma.



This simplified diagram of a chloroplast shows the outer membrane, inner membrane, thylakoids, grana, and stroma.

The chloroplasts contain a green pigment called chlorophyll, which captures the energy of sunlight for photosynthesis. Like plant cells, photosynthetic protists also have chloroplasts. Some bacteria also perform photosynthesis, but they do not have chloroplasts. Their photosynthetic pigments are located in the thylakoid membrane within the cell itself.

Note:
Evolution in Action
Endosymbiosis
We have mentioned that both mitochondria and chloroplasts contain DNA and ribosomes. Have you wondered why? Strong evidence points to endosymbiosis as the explanation. Symbiosis is a relationship in which organisms from two separate species live in close association and typically exhibit specific adaptations to each other. Endosymbiosis (*endo-* = within) is a relationship in which one organism lives inside the other. Endosymbiotic relationships abound in nature. Microbes that produce vitamin K live inside the human gut. This relationship is beneficial for us because we are unable to synthesize vitamin K. It is also beneficial for the microbes because they are protected from other organisms and are provided a stable habitat and abundant food by living within the large intestine.

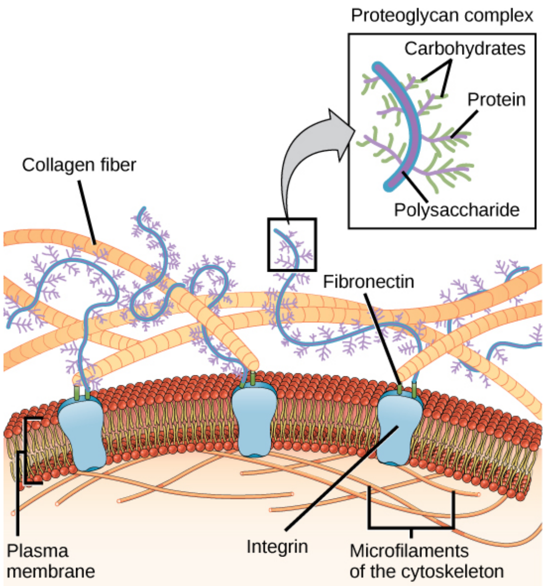
Scientists have long noticed that bacteria, mitochondria, and chloroplasts are similar in size. We also know that mitochondria and chloroplasts have DNA and ribosomes, just as bacteria do. Scientists believe that host cells and bacteria formed a mutually beneficial endosymbiotic relationship when the host cells ingested aerobic bacteria and cyanobacteria but did not destroy them. Through evolution, these ingested bacteria became more specialized in their functions, with the aerobic bacteria becoming mitochondria and the photosynthetic bacteria becoming chloroplasts.

The Central Vacuole

Previously, we mentioned vacuoles as essential components of plant cells. If you look at [link], you will see that plant cells each have a large, central vacuole that occupies most of the cell. The **central vacuole** plays a key role in regulating the cell's concentration of water in changing environmental conditions. In plant cells, the liquid inside the central vacuole provides turgor pressure, which is the outward pressure caused by the fluid inside the cell. Have you ever noticed that if you forget to water a plant for a few days, it wilts? That is because as the water concentration in the soil becomes lower than the water concentration in the plant, water moves out of the central vacuoles and cytoplasm and into the soil. As the central vacuole shrinks, it leaves the cell wall unsupported. This loss of support to the cell walls of a plant results in the wilted appearance. Additionally, this fluid has a very bitter taste, which discourages consumption by insects and animals. The central vacuole also functions to store proteins in developing seed cells.

Extracellular Matrix of Animal Cells

Most animal cells release materials into the extracellular space. The primary components of these materials are glycoproteins and the protein collagen. Collectively, these materials are called the **extracellular matrix** ([link]). Not only does the extracellular matrix hold the cells together to form a tissue, but it also allows the cells within the tissue to communicate with each other.



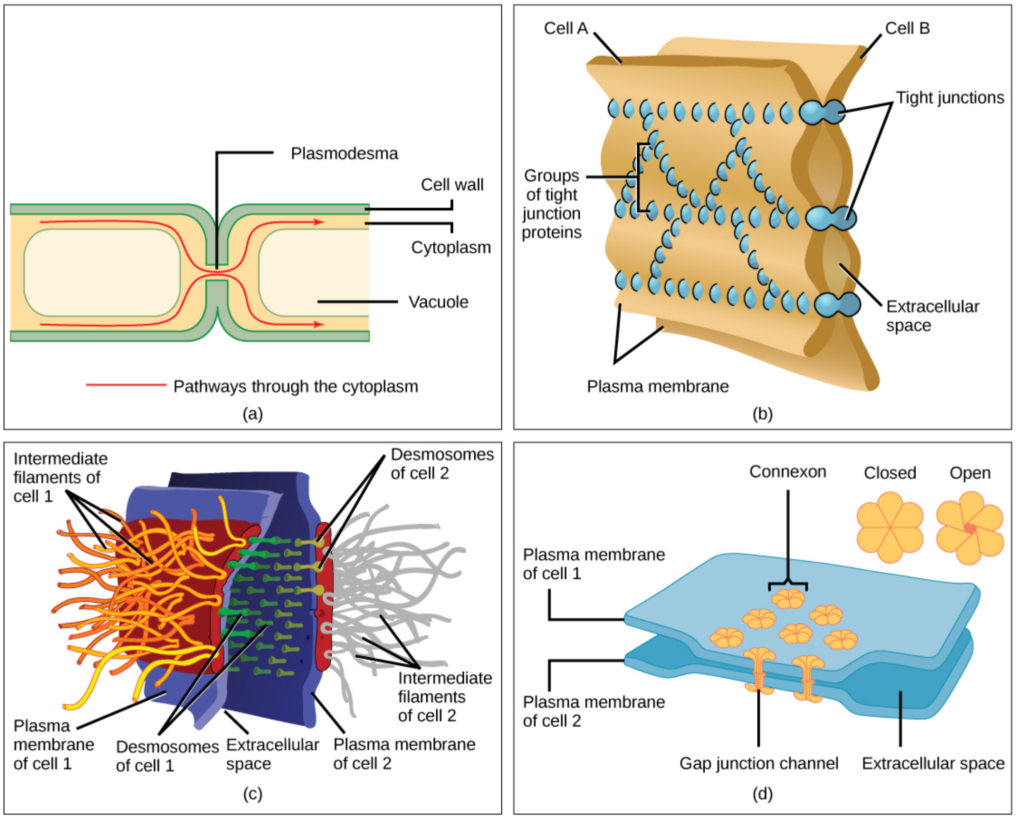
The extracellular matrix consists of a network of substances secreted by cells.

Blood clotting provides an example of the role of the extracellular matrix in cell communication. When the cells lining a blood vessel are damaged, they display a protein receptor called tissue factor. When tissue factor binds with another factor in the extracellular matrix, it causes platelets to adhere to the wall of the damaged blood vessel, stimulates adjacent smooth muscle cells in the blood vessel to contract (thus constricting the blood vessel), and initiates a series of steps that stimulate the platelets to produce clotting factors.

Intercellular Junctions

Cells can also communicate with each other by direct contact, referred to as intercellular junctions. There are some differences in the ways that plant and animal cells do this. **Plasmodesmata** (singular = plasmodesma) are junctions between plant cells, whereas animal cell contacts include tight and gap junctions, and desmosomes.

In general, long stretches of the plasma membranes of neighboring plant cells cannot touch one another because they are separated by the cell walls surrounding each cell. Plasmodesmata are numerous channels that pass between the cell walls of adjacent plant cells, connecting their cytoplasm and enabling signal molecules and nutrients to be transported from cell to cell ([link]**a**).



There are four kinds of connections between cells. (a) A plasmodesma is a channel between the cell walls of two adjacent plant cells. (b) Tight junctions join adjacent animal cells. (c) Desmosomes join two animal cells together. (d) Gap junctions act as channels between animal cells. (credit b, c, d: modification of work by Mariana Ruiz Villareal)

A **tight junction** is a watertight seal between two adjacent animal cells ([link]**b**). Proteins hold the cells tightly against each other. This tight adhesion prevents materials from leaking between the cells. Tight junctions are typically found in the epithelial tissue that lines internal organs and cavities, and composes most of the skin. For example, the tight junctions of the epithelial cells lining the urinary bladder prevent urine from leaking into the extracellular space.

Also found only in animal cells are **desmosomes**, which act like spot welds between adjacent epithelial cells ([link]**c**). They keep cells together in a sheet-like formation in organs and tissues that stretch, like the skin, heart, and muscles.

Gap junctions in animal cells are like plasmodesmata in plant cells in that they are channels between adjacent cells that allow for the transport of ions, nutrients, and other

substances that enable cells to communicate ([link]**d**). Structurally, however, gap junctions and plasmodesmata differ.

Components of Prokaryotic and Eukaryotic Cells and Their Functions				
Cell Component	**Function**	**Present in Prokaryotes?**	**Present in Animal Cells?**	**Present in Plant Cells?**
Plasma membrane	Separates cell from external environment; controls passage of organic molecules, ions, water, oxygen, and wastes into and out of the cell	Yes	Yes	Yes
Cytoplasm	Provides structure to cell; site of many metabolic reactions; medium in which organelles are found	Yes	Yes	Yes
Nucleoid	Location of DNA	Yes	No	No
Nucleus	Cell organelle that houses DNA and directs synthesis of ribosomes and proteins	No	Yes	Yes
Ribosomes	Protein synthesis	Yes	Yes	Yes

Components of Prokaryotic and Eukaryotic Cells and Their Functions				
Cell Component	**Function**	**Present in Prokaryotes?**	**Present in Animal Cells?**	**Present in Plant Cells?**
Mitochondria	ATP production/cellular respiration	No	Yes	Yes
Peroxisomes	Oxidizes and breaks down fatty acids and amino acids, and detoxifies poisons	No	Yes	Yes
Vesicles and vacuoles	Storage and transport; digestive function in plant cells	No	Yes	Yes
Centrosome	Unspecified role in cell division in animal cells; organizing center of microtubules in animal cells	No	Yes	No
Lysosomes	Digestion of macromolecules; recycling of worn-out organelles	No	Yes	No
Cell wall	Protection, structural support and maintenance of cell shape	Yes, primarily peptidoglycan in bacteria but not Archaea	No	Yes, primarily cellulose
Chloroplasts	Photosynthesis	No	No	Yes

Components of Prokaryotic and Eukaryotic Cells and Their Functions				
Cell Component	**Function**	**Present in Prokaryotes?**	**Present in Animal Cells?**	**Present in Plant Cells?**
Endoplasmic reticulum	Modifies proteins and synthesizes lipids	No	Yes	Yes
Golgi apparatus	Modifies, sorts, tags, packages, and distributes lipids and proteins	No	Yes	Yes
Cytoskeleton	Maintains cell’s shape, secures organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move independently	Yes	Yes	Yes
Flagella	Cellular locomotion	Some	Some	No, except for some plant sperm.
Cilia	Cellular locomotion, movement of particles along extracellular surface of plasma membrane, and filtration	No	Some	No

This table provides the components of prokaryotic and eukaryotic cells and their respective functions.

Section Summary

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. The plasma membrane is a phospholipid bilayer embedded with proteins. The nucleolus within the nucleus is the site for ribosome assembly. Ribosomes are found in the cytoplasm or are attached to the cytoplasmic side of the plasma membrane or endoplasmic reticulum. They perform protein synthesis. Mitochondria perform cellular respiration and produce ATP. Peroxisomes break down fatty acids, amino acids, and some toxins. Vesicles and vacuoles are storage and transport compartments. In plant cells, vacuoles also help break down macromolecules.

Animal cells also have a centrosome and lysosomes. The centrosome has two bodies, the centrioles, with an unknown role in cell division. Lysosomes are the digestive organelles of animal cells.

Plant cells have a cell wall, chloroplasts, and a central vacuole. The plant cell wall, whose primary component is cellulose, protects the cell, provides structural support, and gives shape to the cell. Photosynthesis takes place in chloroplasts. The central vacuole expands, enlarging the cell without the need to produce more cytoplasm.

The endomembrane system includes the nuclear envelope, the endoplasmic reticulum, Golgi apparatus, lysosomes, vesicles, as well as the plasma membrane. These cellular components work together to modify, package, tag, and transport membrane lipids and proteins.

The cytoskeleton has three different types of protein elements. Microfilaments provide rigidity and shape to the cell, and facilitate cellular movements. Intermediate filaments bear tension and anchor the nucleus and other organelles in place. Microtubules help the cell resist compression, serve as tracks for motor proteins that move vesicles through the cell, and pull replicated chromosomes to opposite ends of a dividing cell. They are also the structural elements of centrioles, flagella, and cilia.

Animal cells communicate through their extracellular matrices and are connected to each other by tight junctions, desmosomes, and gap junctions. Plant cells are connected and communicate with each other by plasmodesmata.

Art Connections

Exercise:

Problem:

[link] What structures does a plant cell have that an animal cell does not have? What structures does an animal cell have that a plant cell does not have?

Solution:

[link] Plant cells have plasmodesmata, a cell wall, a large central vacuole, chloroplasts, and plastids. Animal cells have lysosomes and centrosomes.

Exercise:

Problem: [link] Why does the *cis* face of the Golgi not face the plasma membrane?

Solution:

[link] Because that face receives chemicals from the ER, which is toward the center of the cell.

Multiple Choice

Exercise:

Problem: Which of the following is found both in eukaryotic and prokaryotic cells?

a. nucleus
b. mitochondrion
c. vacuole
d. ribosome

Solution:

D

Exercise:

Problem: Which of the following is not a component of the endomembrane system?

a. mitochondrion
b. Golgi apparatus
c. endoplasmic reticulum
d. lysosome

Solution:

A

Free Response

Exercise:

Problem:

In the context of cell biology, what do we mean by form follows function? What are at least two examples of this concept?

Solution:

“Form follows function” refers to the idea that the function of a body part dictates the form of that body part. As an example, organisms like birds or fish that fly or swim quickly through the air or water have streamlined bodies that reduce drag. At the level of the cell, in tissues involved in secretory functions, such as the salivary glands, the cells have abundant Golgi.

Glossary

cell wall
: a rigid cell covering made of cellulose in plants, peptidoglycan in bacteria, non-peptidoglycan compounds in Archaea, and chitin in fungi that protects the cell, provides structural support, and gives shape to the cell

central vacuole
: a large plant cell organelle that acts as a storage compartment, water reservoir, and site of macromolecule degradation

chloroplast
: a plant cell organelle that carries out photosynthesis

cilium
: (plural: cilia) a short, hair-like structure that extends from the plasma membrane in large numbers and is used to move an entire cell or move substances along the outer surface of the cell

cytoplasm
: the entire region between the plasma membrane and the nuclear envelope, consisting of organelles suspended in the gel-like cytosol, the cytoskeleton, and various chemicals

cytoskeleton
the network of protein fibers that collectively maintains the shape of the cell, secures some organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move

cytosol
the gel-like material of the cytoplasm in which cell structures are suspended

desmosome
a linkage between adjacent epithelial cells that forms when cadherins in the plasma membrane attach to intermediate filaments

endomembrane system
the group of organelles and membranes in eukaryotic cells that work together to modify, package, and transport lipids and proteins

endoplasmic reticulum (ER)
a series of interconnected membranous structures within eukaryotic cells that collectively modify proteins and synthesize lipids

extracellular matrix
the material, primarily collagen, glycoproteins, and proteoglycans, secreted from animal cells that holds cells together as a tissue, allows cells to communicate with each other, and provides mechanical protection and anchoring for cells in the tissue

flagellum
(plural: flagella) the long, hair-like structure that extends from the plasma membrane and is used to move the cell

gap junction
a channel between two adjacent animal cells that allows ions, nutrients, and other low-molecular weight substances to pass between the cells, enabling the cells to communicate

Golgi apparatus
a eukaryotic organelle made up of a series of stacked membranes that sorts, tags, and packages lipids and proteins for distribution

lysosome
an organelle in an animal cell that functions as the cell’s digestive component; it breaks down proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles

mitochondria

(singular: mitochondrion) the cellular organelles responsible for carrying out cellular respiration, resulting in the production of ATP, the cell’s main energy-carrying molecule

nuclear envelope
: the double-membrane structure that constitutes the outermost portion of the nucleus

nucleolus
: the darkly staining body within the nucleus that is responsible for assembling ribosomal subunits

nucleus
: the cell organelle that houses the cell’s DNA and directs the synthesis of ribosomes and proteins

peroxisome
: a small, round organelle that contains hydrogen peroxide, oxidizes fatty acids and amino acids, and detoxifies many poisons

plasma membrane
: a phospholipid bilayer with embedded (integral) or attached (peripheral) proteins that separates the internal contents of the cell from its surrounding environment

plasmodesma
: (plural: plasmodesmata) a channel that passes between the cell walls of adjacent plant cells, connects their cytoplasm, and allows materials to be transported from cell to cell

ribosome
: a cellular structure that carries out protein synthesis

rough endoplasmic reticulum (RER)
: the region of the endoplasmic reticulum that is studded with ribosomes and engages in protein modification

smooth endoplasmic reticulum (SER)
: the region of the endoplasmic reticulum that has few or no ribosomes on its cytoplasmic surface and synthesizes carbohydrates, lipids, and steroid hormones; detoxifies chemicals like pesticides, preservatives, medications, and environmental pollutants, and stores calcium ions

tight junction
: a firm seal between two adjacent animal cells created by protein adherence

vacuole

a membrane-bound sac, somewhat larger than a vesicle, that functions in cellular storage and transport

vesicle
: a small, membrane-bound sac that functions in cellular storage and transport; its membrane is capable of fusing with the plasma membrane and the membranes of the endoplasmic reticulum and Golgi apparatus

The Cell Membrane
By the end of this section, you will be able to:

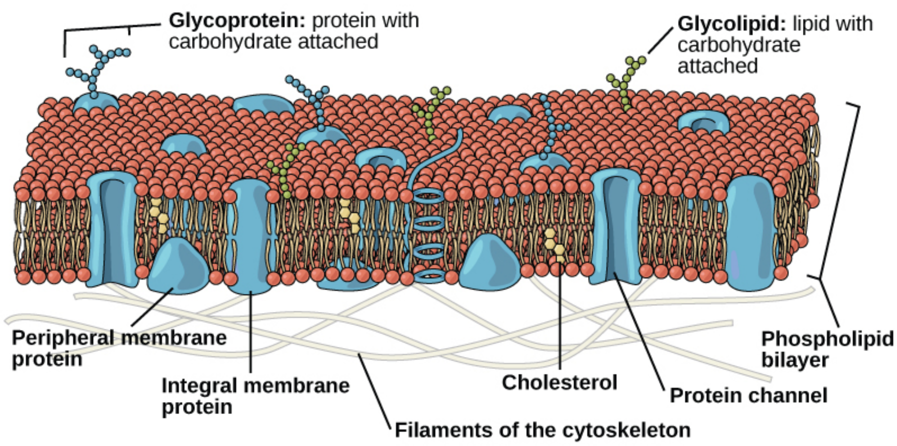
- Understand the fluid mosaic model of membranes
- Describe the functions of phospholipids, proteins, and carbohydrates in membranes

A cell’s plasma membrane defines the boundary of the cell and determines the nature of its contact with the environment. Cells exclude some substances, take in others, and excrete still others, all in controlled quantities. Plasma membranes enclose the borders of cells, but rather than being a static bag, they are dynamic and constantly in flux. The plasma membrane must be sufficiently flexible to allow certain cells, such as red blood cells and white blood cells, to change shape as they pass through narrow capillaries. These are the more obvious functions of a plasma membrane. In addition, the surface of the plasma membrane carries markers that allow cells to recognize one another, which is vital as tissues and organs form during early development, and which later plays a role in the “self” versus “non-self” distinction of the immune response.

The plasma membrane also carries receptors, which are attachment sites for specific substances that interact with the cell. Each receptor is structured to bind with a specific substance. For example, surface receptors of the membrane create changes in the interior, such as changes in enzymes of metabolic pathways. These metabolic pathways might be vital for providing the cell with energy, making specific substances for the cell, or breaking down cellular waste or toxins for disposal. Receptors on the plasma membrane’s exterior surface interact with hormones or neurotransmitters, and allow their messages to be transmitted into the cell. Some recognition sites are used by viruses as attachment points. Although they are highly specific, pathogens like viruses may evolve to exploit receptors to gain entry to a cell by mimicking the specific substance that the receptor is meant to bind. This specificity helps to explain why human immunodeficiency virus (HIV) or any of the five types of hepatitis viruses invade only specific cells.

Fluid Mosaic Model

In 1972, S. J. Singer and Garth L. Nicolson proposed a new model of the plasma membrane that, compared to earlier understanding, better explained both microscopic observations and the function of the plasma membrane. This was called the **fluid mosaic model**. The model has evolved somewhat over time, but still best accounts for the structure and functions of the plasma membrane as we now understand them. The fluid mosaic model describes the structure of the plasma membrane as a mosaic of components—including phospholipids, cholesterol, proteins, and carbohydrates—in which the components are able to flow and change position, while maintaining the basic integrity of the membrane. Both phospholipid molecules and embedded proteins are able to diffuse rapidly and laterally in the membrane. The fluidity of the plasma membrane is necessary for the activities of certain enzymes and transport molecules within the membrane. Plasma membranes range from 5–10 nm thick. As a comparison, human red blood cells, visible via light microscopy, are approximately 8 µm thick, or approximately 1,000 times thicker than a plasma membrane. ([link])



The fluid mosaic model of the plasma membrane structure describes the plasma membrane as a fluid combination of phospholipids, cholesterol, proteins, and carbohydrates.

The plasma membrane is made up primarily of a bilayer of phospholipids with embedded proteins, carbohydrates, glycolipids, and glycoproteins, and, in animal cells, cholesterol. The amount of cholesterol in animal plasma membranes regulates the fluidity of the membrane and changes based on the temperature of the cell's environment. In other words, cholesterol acts as antifreeze in the cell membrane and is more abundant in animals that live in cold climates.

The main fabric of the membrane is composed of two layers of phospholipid molecules, and the polar ends of these molecules (which look like a collection of balls in an artist's rendition of the model) ([link]) are in contact with aqueous fluid both inside and outside the cell. Thus, both surfaces of the plasma membrane are hydrophilic. In contrast, the interior of the membrane, between its two surfaces, is a hydrophobic or nonpolar region because of the fatty acid tails. This region has no attraction for water or other polar molecules.

Proteins make up the second major chemical component of plasma membranes. Integral proteins are embedded in the plasma membrane and may span all or part of the membrane. Integral proteins may serve as channels or pumps to move materials into or out of the cell. Peripheral proteins are found on the exterior or interior surfaces of membranes, attached either to integral proteins or to phospholipid molecules. Both integral and peripheral proteins may serve as enzymes, as structural attachments for the fibers of the cytoskeleton, or as part of the cell's recognition sites.

Carbohydrates are the third major component of plasma membranes. They are always found on the exterior surface of cells and are bound either to proteins (forming glycoproteins) or to lipids (forming glycolipids). These carbohydrate chains may consist of 2–60 monosaccharide units and may be either straight or branched. Along with peripheral proteins, carbohydrates form specialized sites on the cell surface that allow cells to recognize each other.

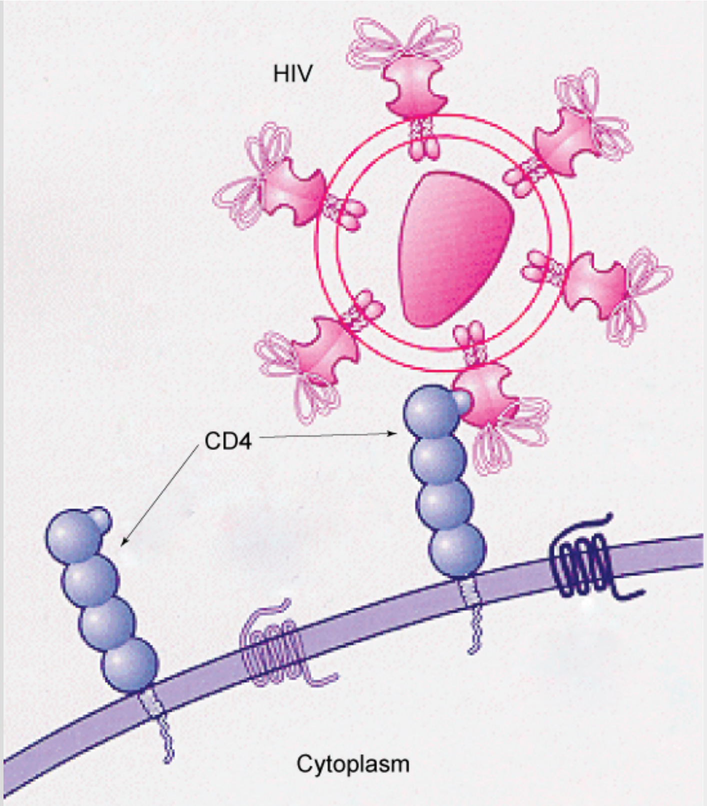
Note:

Evolution in Action

How Viruses Infect Specific Organs

Specific glycoprotein molecules exposed on the surface of the cell membranes of host cells are exploited by many viruses to infect specific organs. For example, HIV is able to penetrate the plasma membranes of specific kinds of white blood cells called T-helper cells and monocytes, as well as some cells of the central nervous system. The hepatitis virus attacks only liver cells.

These viruses are able to invade these cells, because the cells have binding sites on their surfaces that the viruses have exploited with equally specific glycoproteins in their coats. ([link]). The cell is tricked by the mimicry of the virus coat molecules, and the virus is able to enter the cell. Other recognition sites on the virus’s surface interact with the human immune system, prompting the body to produce antibodies. Antibodies are made in response to the antigens (or proteins associated with invasive pathogens). These same sites serve as places for antibodies to attach, and either destroy or inhibit the activity of the virus. Unfortunately, these sites on HIV are encoded by genes that change quickly, making the production of an effective vaccine against the virus very difficult. The virus population within an infected individual quickly evolves through mutation into different populations, or variants, distinguished by differences in these recognition sites. This rapid change of viral surface markers decreases the effectiveness of the person’s immune system in attacking the virus, because the antibodies will not recognize the new variations of the surface patterns.



HIV docks at and binds to the CD4 receptor, a glycoprotein on the surface of T cells, before entering, or infecting, the cell. (credit: modification of work by US National Institutes of Health/National Institute of Allergy and Infectious Diseases)

Section Summary

The modern understanding of the plasma membrane is referred to as the fluid mosaic model. The plasma membrane is composed of a bilayer of phospholipids, with their hydrophobic, fatty acid tails in contact with each other. The landscape of the membrane is studded with proteins, some of

which span the membrane. Some of these proteins serve to transport materials into or out of the cell. Carbohydrates are attached to some of the proteins and lipids on the outward-facing surface of the membrane. These form complexes that function to identify the cell to other cells. The fluid nature of the membrane owes itself to the configuration of the fatty acid tails, the presence of cholesterol embedded in the membrane (in animal cells), and the mosaic nature of the proteins and protein-carbohydrate complexes, which are not firmly fixed in place. Plasma membranes enclose the borders of cells, but rather than being a static bag, they are dynamic and constantly in flux.

Multiple Choice

Exercise:

Problem:

Which plasma membrane component can be either found on its surface or embedded in the membrane structure?

a. protein
b. cholesterol
c. carbohydrate
d. phospholipid

Solution:

A

Exercise:

Problem:

The tails of the phospholipids of the plasma membrane are composed of _____ and are ________?

a. phosphate groups; hydrophobic
b. fatty acid groups; hydrophilic

c. phosphate groups; hydrophilic
d. fatty acid groups; hydrophobic

Solution:

D

Free Response

Exercise:

Problem:

Why is it advantageous for the cell membrane to be fluid in nature?

Solution:

The fluidity of the cell membrane is necessary for the operation of some enzymes and transport mechanisms within the membrane.

Glossary

fluid mosaic model
: a model of the structure of the plasma membrane as a mosaic of components, including phospholipids, cholesterol, proteins, and glycolipids, resulting in a fluid rather than static character

Passive Transport
By the end of this section, you will be able to:

- Explain why and how passive transport occurs
- Understand the processes of osmosis and diffusion
- Define tonicity and describe its relevance to passive transport

Plasma membranes must allow certain substances to enter and leave a cell, while preventing harmful material from entering and essential material from leaving. In other words, plasma membranes are **selectively permeable**—they allow some substances through but not others. If they were to lose this selectivity, the cell would no longer be able to sustain itself, and it would be destroyed. Some cells require larger amounts of specific substances than do other cells; they must have a way of obtaining these materials from the extracellular fluids. This may happen passively, as certain materials move back and forth, or the cell may have special mechanisms that ensure transport. Most cells expend most of their energy, in the form of adenosine triphosphate (ATP), to create and maintain an uneven distribution of ions on the opposite sides of their membranes. The structure of the plasma membrane contributes to these functions, but it also presents some problems.

The most direct forms of membrane transport are passive. **Passive transport** is a naturally occurring phenomenon and does not require the cell to expend energy to accomplish the movement. In passive transport, substances move from an area of higher concentration to an area of lower concentration in a process called diffusion. A physical space in which there is a different concentration of a single substance is said to have a **concentration gradient**.

Selective Permeability

Plasma membranes are asymmetric, meaning that despite the mirror image formed by the phospholipids, the interior of the membrane is not identical to the exterior of the membrane. Integral proteins that act as channels or pumps work in one direction. Carbohydrates, attached to lipids or proteins, are also found on the exterior surface of the plasma membrane. These

carbohydrate complexes help the cell bind substances that the cell needs in the extracellular fluid. This adds considerably to the selective nature of plasma membranes.

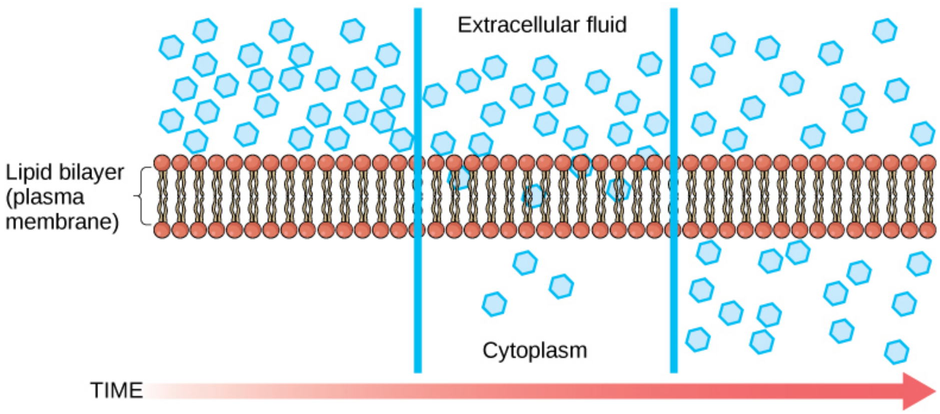
Recall that plasma membranes have hydrophilic and hydrophobic regions. This characteristic helps the movement of certain materials through the membrane and hinders the movement of others. Lipid-soluble material can easily slip through the hydrophobic lipid core of the membrane. Substances such as the fat-soluble vitamins A, D, E, and K readily pass through the plasma membranes in the digestive tract and other tissues. Fat-soluble drugs also gain easy entry into cells and are readily transported into the body's tissues and organs. Molecules of oxygen and carbon dioxide have no charge and pass through by simple diffusion.

Polar substances, with the exception of water, present problems for the membrane. While some polar molecules connect easily with the outside of a cell, they cannot readily pass through the lipid core of the plasma membrane. Additionally, whereas small ions could easily slip through the spaces in the mosaic of the membrane, their charge prevents them from doing so. Ions such as sodium, potassium, calcium, and chloride must have a special means of penetrating plasma membranes. Simple sugars and amino acids also need help with transport across plasma membranes.

Diffusion

Diffusion is a passive process of transport. A single substance tends to move from an area of high concentration to an area of low concentration until the concentration is equal across the space. You are familiar with diffusion of substances through the air. For example, think about someone opening a bottle of perfume in a room filled with people. The perfume is at its highest concentration in the bottle and is at its lowest at the edges of the room. The perfume vapor will diffuse, or spread away, from the bottle, and gradually, more and more people will smell the perfume as it spreads. Materials move within the cell's cytosol by diffusion, and certain materials move through the plasma membrane by diffusion ([link]). Diffusion expends no energy. Rather the different concentrations of materials in different areas are a form of potential energy, and diffusion is the

dissipation of that potential energy as materials move down their concentration gradients, from high to low.



Diffusion through a permeable membrane follows the concentration gradient of a substance, moving the substance from an area of high concentration to one of low concentration. (credit: modification of work by Mariana Ruiz Villarreal)

Each separate substance in a medium, such as the extracellular fluid, has its own concentration gradient, independent of the concentration gradients of other materials. Additionally, each substance will diffuse according to that gradient.

Several factors affect the rate of diffusion.

- Extent of the concentration gradient: The greater the difference in concentration, the more rapid the diffusion. The closer the distribution of the material gets to equilibrium, the slower the rate of diffusion becomes.
- Mass of the molecules diffusing: More massive molecules move more slowly, because it is more difficult for them to move between the molecules of the substance they are moving through; therefore, they diffuse more slowly.

- Temperature: Higher temperatures increase the energy and therefore the movement of the molecules, increasing the rate of diffusion.
- Solvent density: As the density of the solvent increases, the rate of diffusion decreases. The molecules slow down because they have a more difficult time getting through the denser medium.

Note:
Concept in Action

For an animation of the diffusion process in action, view this short video on cell membrane transport.

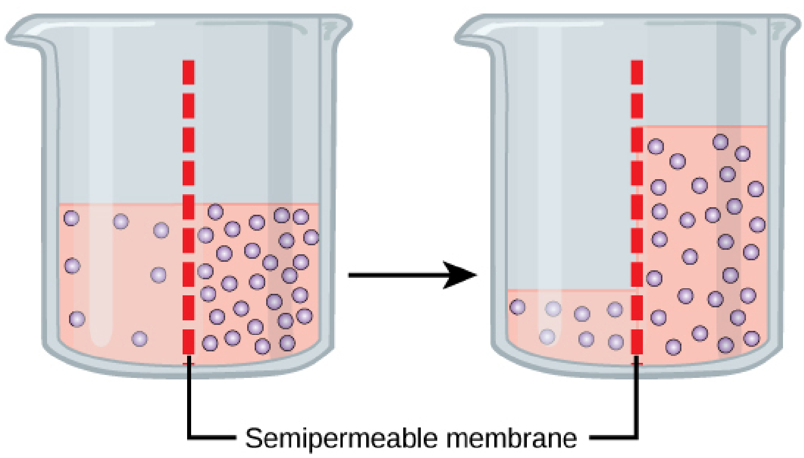
Facilitated transport

In **facilitated transport**, also called facilitated diffusion, material moves across the plasma membrane with the assistance of transmembrane proteins down a concentration gradient (from high to low concentration) without the expenditure of cellular energy. However, the substances that undergo facilitated transport would otherwise not diffuse easily or quickly across the plasma membrane. The solution to moving polar substances and other substances across the plasma membrane rests in the proteins that span its surface. The material being transported is first attached to protein or glycoprotein receptors on the exterior surface of the plasma membrane. This allows the material that is needed by the cell to be removed from the extracellular fluid. The substances are then passed to specific integral proteins that facilitate their passage, because they form channels or pores that allow certain substances to pass through the membrane. The integral proteins involved in facilitated transport are collectively referred to as

transport proteins, and they function as either channels for the material or carriers.

Osmosis

Osmosis is the diffusion of water through a semipermeable membrane according to the concentration gradient of water across the membrane. Whereas diffusion transports material across membranes and within cells, osmosis transports *only water* across a membrane and the membrane limits the diffusion of solutes in the water. Osmosis is a special case of diffusion. Water, like other substances, moves from an area of higher concentration to one of lower concentration. Imagine a beaker with a semipermeable membrane, separating the two sides or halves ([link]). On both sides of the membrane, the water level is the same, but there are different concentrations on each side of a dissolved substance, or **solute**, that cannot cross the membrane. If the volume of the water is the same, but the concentrations of solute are different, then there are also different concentrations of water, the solvent, on either side of the membrane.



In osmosis, water always moves from an area of higher concentration (of water) to one of lower concentration (of water). In this system, the solute cannot pass through the selectively permeable membrane.

A principle of diffusion is that the molecules move around and will spread evenly throughout the medium if they can. However, only the material capable of getting through the membrane will diffuse through it. In this example, the solute cannot diffuse through the membrane, but the water can. Water has a concentration gradient in this system. Therefore, water will diffuse down its concentration gradient, crossing the membrane to the side where it is less concentrated. This diffusion of water through the membrane—osmosis—will continue until the concentration gradient of water goes to zero. Osmosis proceeds constantly in living systems.

Note:
Concept in Action

Watch this video that illustrates diffusion in hot versus cold solutions.

Tonicity

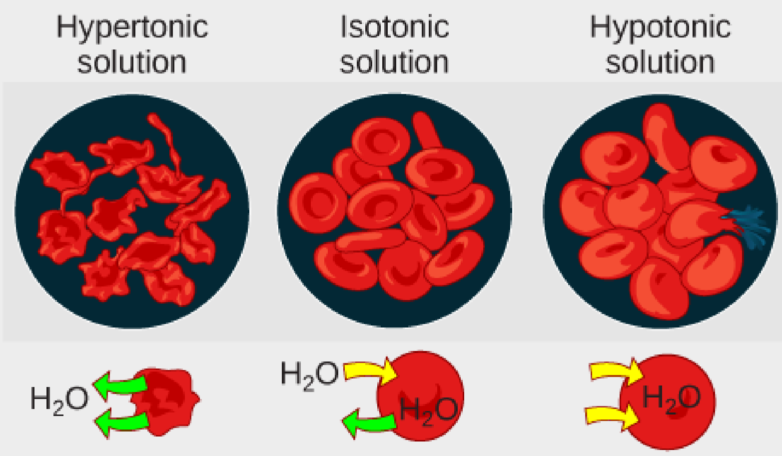
Tonicity describes the amount of solute in a solution. The measure of the tonicity of a solution, or the total amount of solutes dissolved in a specific amount of solution, is called its **osmolarity**. Three terms—hypotonic, isotonic, and hypertonic—are used to relate the osmolarity of a cell to the osmolarity of the extracellular fluid that contains the cells. In a **hypotonic** solution, such as tap water, the extracellular fluid has a lower concentration of solutes than the fluid inside the cell, and water enters the cell. (In living systems, the point of reference is always the cytoplasm, so the prefix *hypo-*

means that the extracellular fluid has a lower concentration of solutes, or a lower osmolarity, than the cell cytoplasm.) It also means that the extracellular fluid has a higher concentration of water than does the cell. In this situation, water will follow its concentration gradient and enter the cell. This may cause an animal cell to burst, or lyse.

In a **hypertonic** solution (the prefix *hyper*- refers to the extracellular fluid having a higher concentration of solutes than the cell's cytoplasm), the fluid contains less water than the cell does, such as seawater. Because the cell has a lower concentration of solutes, the water will leave the cell. In effect, the solute is drawing the water out of the cell. This may cause an animal cell to shrivel, or crenate.

In an **isotonic** solution, the extracellular fluid has the same osmolarity as the cell. If the concentration of solutes of the cell matches that of the extracellular fluid, there will be no net movement of water into or out of the cell. Blood cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances ([link]).

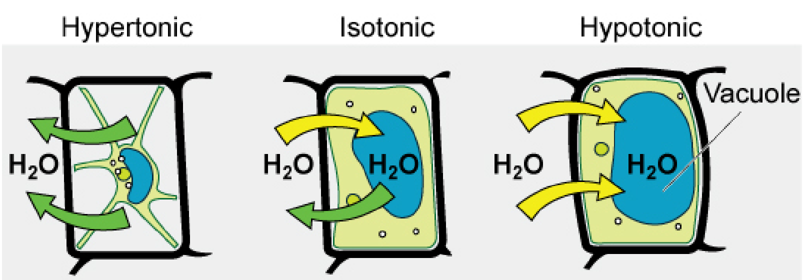
Note:
Art Connection



Osmotic pressure changes the shape of red blood cells in hypertonic, isotonic, and hypotonic solutions. (credit: modification of work by Mariana Ruiz Villarreal)

A doctor injects a patient with what the doctor thinks is isotonic saline solution. The patient dies, and autopsy reveals that many red blood cells have been destroyed. Do you think the solution the doctor injected was really isotonic?

Some organisms, such as plants, fungi, bacteria, and some protists, have cell walls that surround the plasma membrane and prevent cell lysis. The plasma membrane can only expand to the limit of the cell wall, so the cell will not lyse. In fact, the cytoplasm in plants is always slightly hypertonic compared to the cellular environment, and water will always enter a cell if water is available. This influx of water produces turgor pressure, which stiffens the cell walls of the plant ([link]). In nonwoody plants, turgor pressure supports the plant. If the plant cells become hypertonic, as occurs in drought or if a plant is not watered adequately, water will leave the cell. Plants lose turgor pressure in this condition and wilt.



The turgor pressure within a plant cell depends on the tonicity of the solution that it is bathed in. (credit: modification of work by Mariana Ruiz Villarreal)

Section Summary

The passive forms of transport, diffusion and osmosis, move material of small molecular weight. Substances diffuse from areas of high concentration to areas of low concentration, and this process continues until the substance is evenly distributed in a system. In solutions of more than one substance, each type of molecule diffuses according to its own concentration gradient. Many factors can affect the rate of diffusion, including concentration gradient, the sizes of the particles that are diffusing, and the temperature of the system.

In living systems, diffusion of substances into and out of cells is mediated by the plasma membrane. Some materials diffuse readily through the membrane, but others are hindered, and their passage is only made possible by protein channels and carriers. The chemistry of living things occurs in aqueous solutions, and balancing the concentrations of those solutions is an ongoing problem. In living systems, diffusion of some substances would be slow or difficult without membrane proteins.

Art Connections

Exercise:

Problem:

[link] A doctor injects a patient with what he thinks is isotonic saline solution. The patient dies, and autopsy reveals that many red blood cells have been destroyed. Do you think the solution the doctor injected was really isotonic?

Solution:

[link] No, it must have been hypotonic, as a hypotonic solution would cause water to enter the cells, thereby making them burst.

Multiple Choice

Exercise:

Problem: Water moves via osmosis _________.

a. throughout the cytoplasm
b. from an area with a high concentration of other solutes to a lower one
c. from an area with a low concentration of solutes to an area with a higher one
d. from an area with a low concentration of water to one of higher concentration

---

Solution:

C

Exercise:

Problem:

The principal force driving movement in diffusion is __________.

a. temperature
b. particle size
c. concentration gradient
d. membrane surface area

Free Response

Exercise:

Problem: Why does osmosis occur?

Solution:

Water moves through a semipermeable membrane in osmosis because there is a concentration gradient across the membrane of solute and

solvent. The solute cannot effectively move to balance the concentration on both sides of the membrane, so water moves to achieve this balance.

Glossary

concentration gradient
: an area of high concentration across from an area of low concentration

diffusion
: a passive process of transport of low-molecular weight material down its concentration gradient

facilitated transport
: a process by which material moves down a concentration gradient (from high to low concentration) using integral membrane proteins

hypertonic
: describes a solution in which extracellular fluid has higher osmolarity than the fluid inside the cell

hypotonic
: describes a solution in which extracellular fluid has lower osmolarity than the fluid inside the cell

isotonic
: describes a solution in which the extracellular fluid has the same osmolarity as the fluid inside the cell

osmolarity
: the total amount of substances dissolved in a specific amount of solution

osmosis
: the transport of water through a semipermeable membrane from an area of high water concentration to an area of low water concentration across a membrane

passive transport
: a method of transporting material that does not require energy

selectively permeable
: the characteristic of a membrane that allows some substances through but not others

solute
: a substance dissolved in another to form a solution

tonicity
: the amount of solute in a solution.

Active Transport
By the end of this section, you will be able to:

- Understand how electrochemical gradients affect ions
- Describe endocytosis, including phagocytosis, pinocytosis, and receptor-mediated endocytosis
- Understand the process of exocytosis

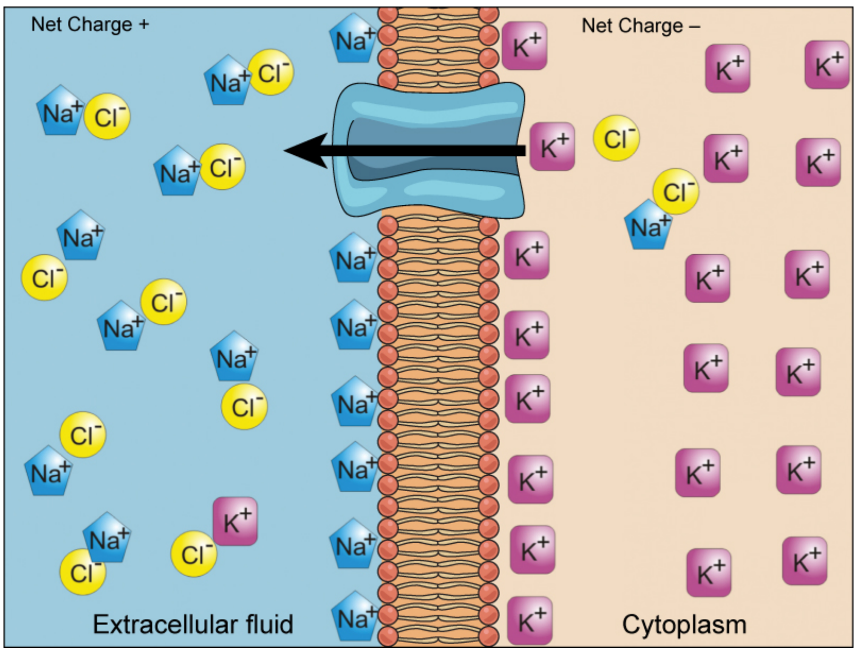
Active transport mechanisms require the use of the cell's energy, usually in the form of adenosine triphosphate (ATP). If a substance must move into the cell against its concentration gradient, that is, if the concentration of the substance inside the cell must be greater than its concentration in the extracellular fluid, the cell must use energy to move the substance. Some active transport mechanisms move small-molecular weight material, such as ions, through the membrane.

In addition to moving small ions and molecules through the membrane, cells also need to remove and take in larger molecules and particles. Some cells are even capable of engulfing entire unicellular microorganisms. You might have correctly hypothesized that the uptake and release of large particles by the cell requires energy. A large particle, however, cannot pass through the membrane, even with energy supplied by the cell.

Electrochemical Gradient

We have discussed simple concentration gradients—differential concentrations of a substance across a space or a membrane—but in living systems, gradients are more complex. Because cells contain proteins, most of which are negatively charged, and because ions move into and out of cells, there is an electrical gradient, a difference of charge, across the plasma membrane. The interior of living cells is electrically negative with respect to the extracellular fluid in which they are bathed; at the same time, cells have higher concentrations of potassium (K^+) and lower concentrations of sodium (Na^+) than does the extracellular fluid. Thus, in a living cell, the concentration gradient and electrical gradient of Na^+ promotes diffusion of the ion into the cell, and the electrical gradient of Na^+ (a positive ion) tends to drive it inward to the negatively charged interior.

The situation is more complex, however, for other elements such as potassium. The electrical gradient of K^+ promotes diffusion of the ion *into* the cell, but the concentration gradient of K^+ promotes diffusion *out* of the cell ([link]). The combined gradient that affects an ion is called its **electrochemical gradient**, and it is especially important to muscle and nerve cells.



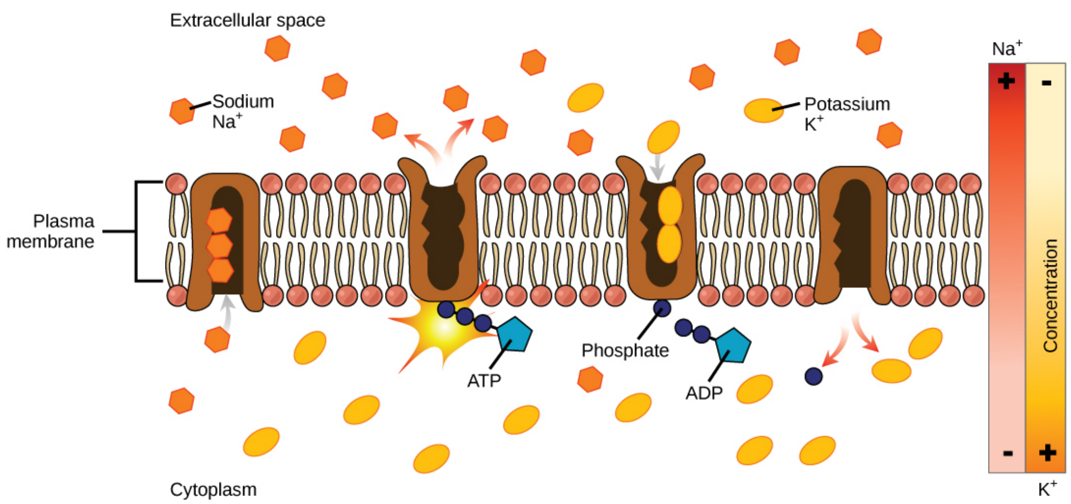
Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients. (credit: modification of work by “Synaptitude”/Wikimedia Commons)

Moving Against a Gradient

To move substances against a concentration or an electrochemical gradient, the cell must use energy. This energy is harvested from ATP that is generated through cellular metabolism. Active transport mechanisms,

collectively called pumps or carrier proteins, work against electrochemical gradients. With the exception of ions, small substances constantly pass through plasma membranes. Active transport maintains concentrations of ions and other substances needed by living cells in the face of these passive changes. Much of a cell's supply of metabolic energy may be spent maintaining these processes. Because active transport mechanisms depend on cellular metabolism for energy, they are sensitive to many metabolic poisons that interfere with the supply of ATP.

Two mechanisms exist for the transport of small-molecular weight material and macromolecules. Primary active transport moves ions across a membrane and creates a difference in charge across that membrane. The primary active transport system uses ATP to move a substance, such as an ion, into the cell, and often at the same time, a second substance is moved out of the cell. The sodium-potassium pump, an important pump in animal cells, expends energy to move potassium ions into the cell and a different number of sodium ions out of the cell ([link]). The action of this pump results in a concentration and charge difference across the membrane.

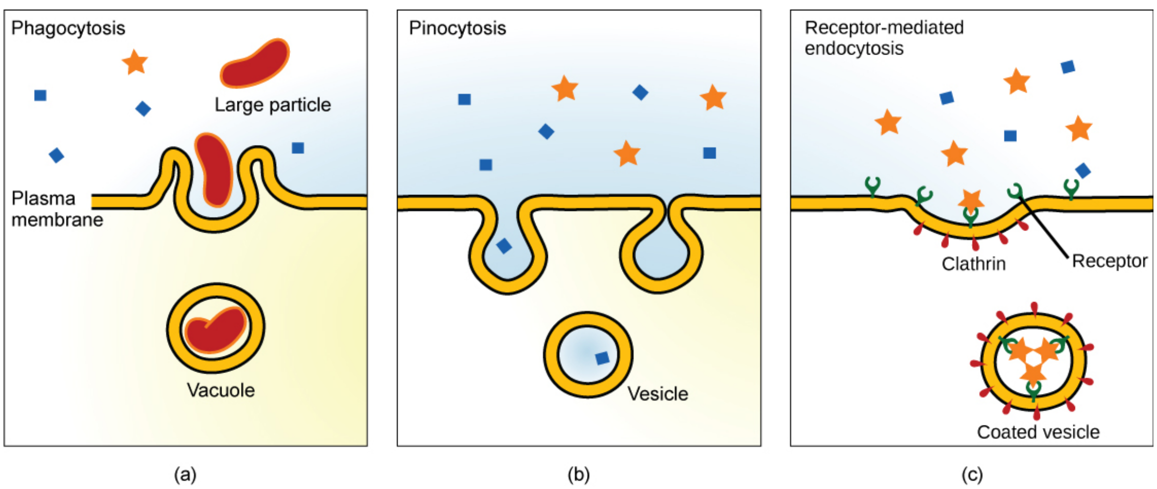


The sodium-potassium pump move potassium and sodium ions across the plasma membrane. (credit: modification of work by Mariana Ruiz Villarreal)

Secondary active transport describes the movement of material using the energy of the electrochemical gradient established by primary active transport. Using the energy of the electrochemical gradient created by the primary active transport system, other substances such as amino acids and glucose can be brought into the cell through membrane channels. ATP itself is formed through secondary active transport using a hydrogen ion gradient in the mitochondrion.

Endocytosis

Endocytosis is a type of active transport that moves particles, such as large molecules, parts of cells, and even whole cells, into a cell. There are different variations of endocytosis, but all share a common characteristic: The plasma membrane of the cell invaginates, forming a pocket around the target particle. The pocket pinches off, resulting in the particle being contained in a newly created vacuole that is formed from the plasma membrane.



Three variations of endocytosis are shown. (a) In one form of endocytosis, phagocytosis, the cell membrane surrounds the particle and pinches off to form an intracellular vacuole. (b) In another type of endocytosis, pinocytosis, the cell membrane surrounds a small volume of fluid and pinches off, forming a

vesicle. (c) In receptor-mediated endocytosis, uptake of substances by the cell is targeted to a single type of substance that binds at the receptor on the external cell membrane. (credit: modification of work by Mariana Ruiz Villarreal)

Phagocytosis is the process by which large particles, such as cells, are taken in by a cell. For example, when microorganisms invade the human body, a type of white blood cell called a neutrophil removes the invader through this process, surrounding and engulfing the microorganism, which is then destroyed by the neutrophil ([link]).

A variation of endocytosis is called **pinocytosis**. This literally means “cell drinking” and was named at a time when the assumption was that the cell was purposefully taking in extracellular fluid. In reality, this process takes in solutes that the cell needs from the extracellular fluid ([link]).

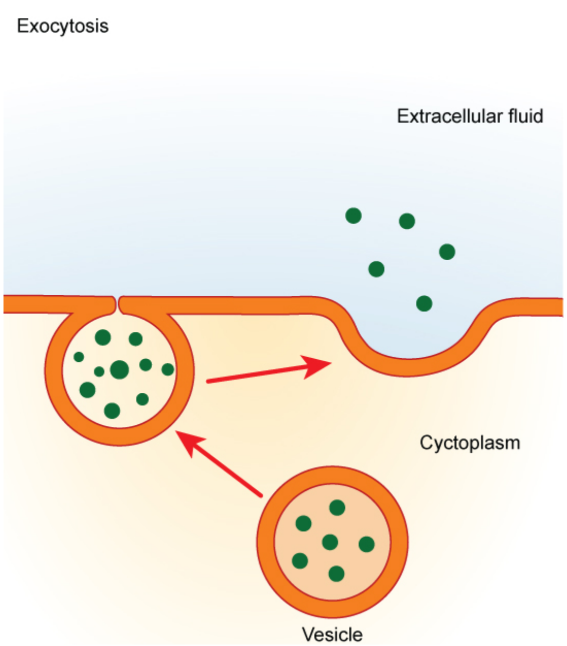
A targeted variation of endocytosis employs binding proteins in the plasma membrane that are specific for certain substances ([link]). The particles bind to the proteins and the plasma membrane invaginates, bringing the substance and the proteins into the cell. If passage across the membrane of the target of **receptor-mediated endocytosis** is ineffective, it will not be removed from the tissue fluids or blood. Instead, it will stay in those fluids and increase in concentration. Some human diseases are caused by a failure of receptor-mediated endocytosis. For example, the form of cholesterol termed low-density lipoprotein or LDL (also referred to as “bad” cholesterol) is removed from the blood by receptor-mediated endocytosis. In the human genetic disease familial hypercholesterolemia, the LDL receptors are defective or missing entirely. People with this condition have life-threatening levels of cholesterol in their blood, because their cells cannot clear the chemical from their blood.

Note:
Concept in Action

See receptor-mediated endocytosis [animation](#) in action.

Exocytosis

In contrast to these methods of moving material into a cell is the process of exocytosis. **Exocytosis** is the opposite of the processes discussed above in that its purpose is to expel material from the cell into the extracellular fluid. A particle enveloped in membrane fuses with the interior of the plasma membrane. This fusion opens the membranous envelope to the exterior of the cell, and the particle is expelled into the extracellular space ([link]).



In exocytosis, a vesicle migrates to the plasma

membrane, binds, and releases its contents to the outside of the cell. (credit: modification of work by Mariana Ruiz Villarreal)

Section Summary

The combined gradient that affects an ion includes its concentration gradient and its electrical gradient. Living cells need certain substances in concentrations greater than they exist in the extracellular space. Moving substances up their electrochemical gradients requires energy from the cell. Active transport uses energy stored in ATP to fuel the transport. Active transport of small molecular-size material uses integral proteins in the cell membrane to move the material—these proteins are analogous to pumps. Some pumps, which carry out primary active transport, couple directly with ATP to drive their action. In secondary transport, energy from primary transport can be used to move another substance into the cell and up its concentration gradient.

Endocytosis methods require the direct use of ATP to fuel the transport of large particles such as macromolecules; parts of cells or whole cells can be engulfed by other cells in a process called phagocytosis. In phagocytosis, a portion of the membrane invaginates and flows around the particle, eventually pinching off and leaving the particle wholly enclosed by an envelope of plasma membrane. Vacuoles are broken down by the cell, with the particles used as food or dispatched in some other way. Pinocytosis is a similar process on a smaller scale. The cell expels waste and other particles through the reverse process, exocytosis. Wastes are moved outside the cell, pushing a membranous vesicle to the plasma membrane, allowing the vesicle to fuse with the membrane and incorporating itself into the membrane structure, releasing its contents to the exterior of the cell.

Multiple Choice

Exercise:

Problem:

Active transport must function continuously because ___________.

a. plasma membranes wear out
b. cells must be in constant motion
c. facilitated transport opposes active transport
d. diffusion is constantly moving the solutes in the other direction

Solution:

D

Free Response

Exercise:

Problem:

Where does the cell get energy for active transport processes?

Solution:

The cell harvests energy from ATP produced by its own metabolism to power active transport processes, such as pumps.

Glossary

active transport
: the method of transporting material that requires energy

electrochemical gradient
: a gradient produced by the combined forces of the electrical gradient and the chemical gradient

endocytosis

a type of active transport that moves substances, including fluids and particles, into a cell

exocytosis

a process of passing material out of a cell

phagocytosis

a process that takes macromolecules that the cell needs from the extracellular fluid; a variation of endocytosis

pinocytosis

a process that takes solutes that the cell needs from the extracellular fluid; a variation of endocytosis

receptor-mediated endocytosis

a variant of endocytosis that involves the use of specific binding proteins in the plasma membrane for specific molecules or particles

Introduction class="introduction"

A hummingbird needs energy to maintain prolonged flight. The bird obtains its energy from taking in food and transforming the energy contained in food molecules into forms of energy to power its flight through a series of biochemical reactions. (credit: modification of work by Cory Zanker)

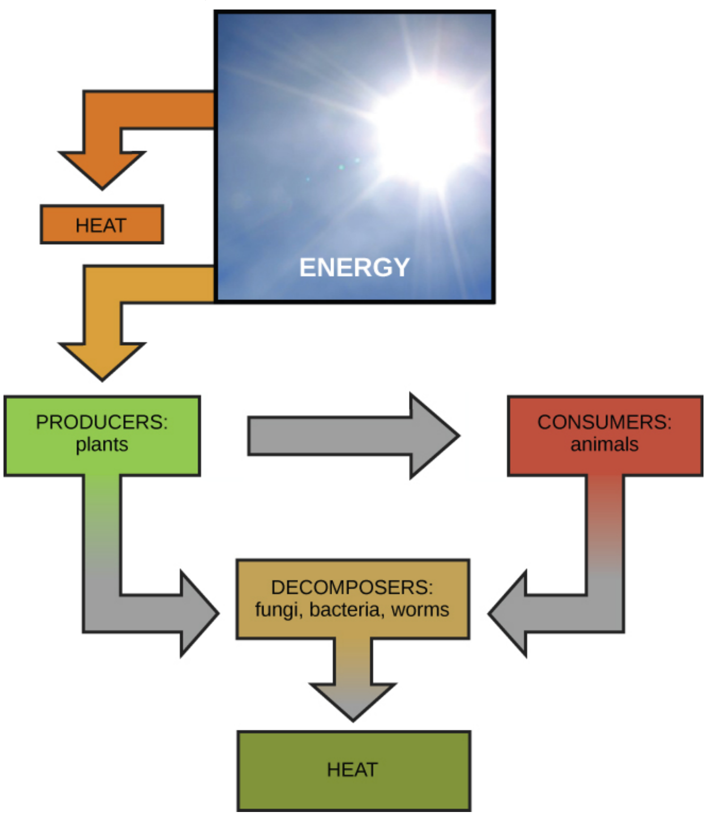
Virtually every task performed by living organisms requires energy. Energy is needed to perform heavy labor and exercise, but humans also use energy while thinking, and even during sleep. In fact, the living cells of every organism constantly use energy. Nutrients and other molecules are imported into the cell, metabolized (broken down) and possibly synthesized into new molecules, modified if needed, transported around the cell, and possibly distributed to the entire organism. For example, the large proteins that make up muscles are built from smaller molecules imported from dietary amino acids. Complex carbohydrates are broken down into simple sugars that the cell uses for energy. Just as energy is required to both build and demolish a building, energy is required for the synthesis and breakdown of molecules as well as the transport of molecules into and out of cells. In addition, processes such as ingesting and breaking down pathogenic bacteria and viruses, exporting wastes and toxins, and movement of the cell require energy. From where, and in what form, does this energy come? How do living cells obtain energy, and how do they use it? This chapter will discuss different forms of energy and the physical laws that govern energy transfer. This chapter will also describe how cells use energy and replenish it, and how chemical reactions in the cell are performed with great efficiency.

Energy and Metabolism

By the end of this section, you will be able to:

- Explain what metabolic pathways are
- State the first and second laws of thermodynamics
- Explain the difference between kinetic and potential energy
- Describe endergonic and exergonic reactions
- Discuss how enzymes function as molecular catalysts

Scientists use the term **bioenergetics** to describe the concept of energy flow ([link]) through living systems, such as cells. Cellular processes such as the building and breaking down of complex molecules occur through stepwise chemical reactions. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. Just as living things must continually consume food to replenish their energy supplies, cells must continually produce more energy to replenish that used by the many energy-requiring chemical reactions that constantly take place. Together, all of the chemical reactions that take place inside cells, including those that consume or generate energy, are referred to as the cell’s **metabolism**.



Ultimately, most life forms get their energy from the sun. Plants use photosynthesis to capture sunlight, and herbivores eat the plants to obtain energy. Carnivores eat the herbivores, and eventual decomposition of plant and animal material contributes to the nutrient pool.

Metabolic Pathways

Consider the metabolism of sugar. This is a classic example of one of the many cellular processes that use and produce energy. Living things consume sugars as a major energy source, because sugar molecules have a great deal of energy stored within their bonds. For the most part, photosynthesizing organisms like plants produce these sugars. During photosynthesis, plants use energy (originally from sunlight) to convert carbon dioxide gas (CO_2) into sugar molecules (like glucose: $C_6H_{12}O_6$). They consume carbon dioxide and produce oxygen as a waste product. This reaction is summarized as:

Equation:

$$6CO_2 + 6H_2O \text{ --> } C_6H_{12}O_6 + 6O_2$$

Because this process involves synthesizing an energy-storing molecule, it requires energy input to proceed. During the light reactions of photosynthesis, energy is provided by a molecule called adenosine triphosphate (ATP), which is the primary energy currency of all cells. Just as the dollar is used as currency to buy goods, cells use molecules of ATP as energy currency to perform immediate work. In contrast, energy-storage molecules such as glucose are consumed only to be broken down to use their energy. The reaction that harvests the energy of a sugar molecule in cells requiring oxygen to survive can be summarized by the reverse reaction

to photosynthesis. In this reaction, oxygen is consumed and carbon dioxide is released as a waste product. The reaction is summarized as:

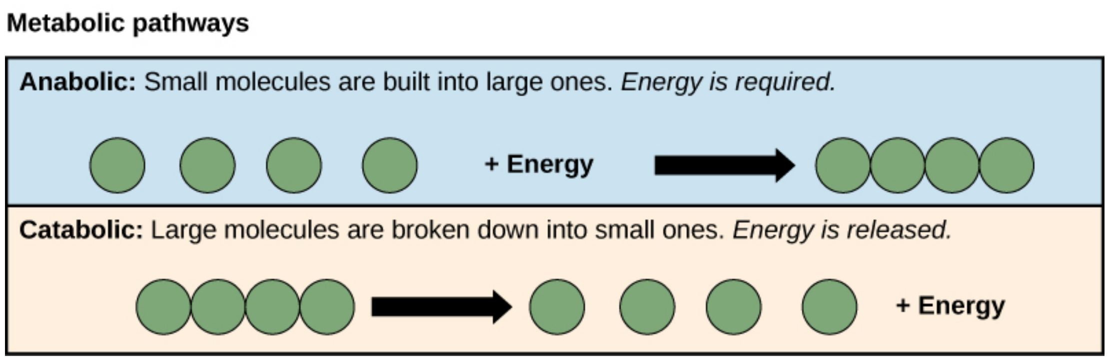
Equation:

$$C_6H_{12}O_6 + 6O_2 \text{-->} 6H_2O + 6CO_2$$

Both of these reactions involve many steps.

The processes of making and breaking down sugar molecules illustrate two examples of metabolic pathways. A metabolic pathway is a series of chemical reactions that takes a starting molecule and modifies it, step-by-step, through a series of metabolic intermediates, eventually yielding a final product. In the example of sugar metabolism, the first metabolic pathway synthesized sugar from smaller molecules, and the other pathway broke sugar down into smaller molecules. These two opposite processes—the first requiring energy and the second producing energy—are referred to as **anabolic** pathways (building polymers) and **catabolic** pathways (breaking down polymers into their monomers), respectively. Consequently, metabolism is composed of synthesis (anabolism) and degradation (catabolism) ([link]).

It is important to know that the chemical reactions of metabolic pathways do not take place on their own. Each reaction step is facilitated, or catalyzed, by a protein called an enzyme. Enzymes are important for catalyzing all types of biological reactions—those that require energy as well as those that release energy.



Catabolic pathways are those that generate energy by breaking down larger molecules. Anabolic pathways are those that require energy to synthesize larger molecules. Both types of pathways are required for maintaining the cell’s energy balance.

Energy

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter relevant to a particular case of energy transfer is called a system, and everything outside of that matter is called the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy is transferred within the system (between the stove, pot, and water). There are two types of systems: open and closed. In an open system, energy can be exchanged with its surroundings. The stovetop system is open because heat can be lost to the air. A closed system cannot exchange energy with its surroundings.

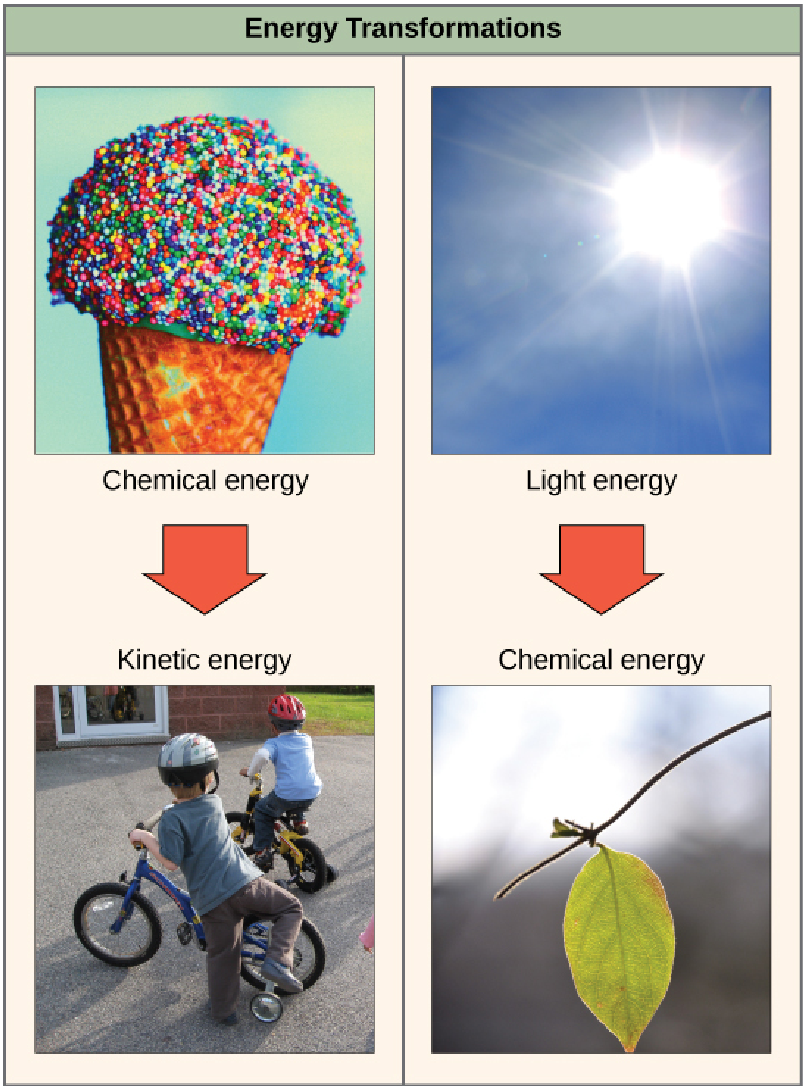
Biological organisms are open systems. Energy is exchanged between them and their surroundings as they use energy from the sun to perform photosynthesis or consume energy-storing molecules and release energy to the environment by doing work and releasing heat. Like all things in the physical world, energy is subject to physical laws. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe.

In general, energy is defined as the ability to do work, or to create some kind of change. Energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different types of energy. To appreciate the way energy flows into and out of biological systems, it is important to understand two of the physical laws that govern energy.

Thermodynamics

The first law of thermodynamics states that the total amount of energy in the universe is constant and conserved. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. Energy exists in many different forms. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, but it cannot be created or destroyed. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light and heat energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight to chemical energy stored within organic molecules ([link]). Some examples of energy transformations are shown in [link].

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge. Chemical energy stored within organic molecules such as sugars and fats is transferred and transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the motion of cilia or flagella, and contracting muscle fibers to create movement.



Shown are some examples of energy transferred and transformed from one system to another and from one form to another. The food we consume provides our cells with the energy required to carry out bodily functions, just as light energy provides plants with the means to create the chemical energy they need. (credit "ice cream": modification of work by D. Sharon Pruitt; credit "kids": modification of work by Max from Providence; credit "leaf": modification of work by Cory Zanker)

A living cell's primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the second law of thermodynamics explains why these tasks are harder than they appear. All energy transfers and transformations are never completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy. Thermodynamically, **heat energy** is defined as the energy transferred from one system to another that is not work. For example, when a light bulb is turned on, some of the energy being converted from electrical energy into light energy is lost as heat energy. Likewise, some energy is lost as heat energy during cellular metabolic reactions.

An important concept in physical systems is that of order and disorder. The more energy that is lost by a system to its surroundings, the less ordered and more random the system is. Scientists refer to the measure of randomness or disorder within a system as entropy. High entropy means high disorder and low energy. Molecules and chemical reactions have varying entropy as well. For example, entropy increases as molecules at a high concentration in one place diffuse and spread out. The second law of thermodynamics says that energy will always be lost as heat in energy transfers or transformations.

Living things are highly ordered, requiring constant energy input to be maintained in a state of low entropy.

Potential and Kinetic Energy

When an object is in motion, there is energy associated with that object. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. Energy associated with objects in motion is called **kinetic energy** ([link]). A speeding bullet, a walking person, and the rapid movement of molecules in the air (which produces heat) all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above ground with a crane? If the suspended wrecking ball is unmoving, is there

energy associated with it? The answer is yes. The energy that was required to lift the wrecking ball did not disappear, but is now stored in the wrecking ball by virtue of its position and the force of gravity acting on it. This type of energy is called **potential energy** ([link]). If the ball were to fall, the potential energy would be transformed into kinetic energy until all of the potential energy was exhausted when the ball rested on the ground. Wrecking balls also swing like a pendulum; through the swing, there is a constant change of potential energy (highest at the top of the swing) to kinetic energy (highest at the bottom of the swing). Other examples of potential energy include the energy of water held behind a dam or a person about to skydive out of an airplane.

Still water has potential energy; moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit "dam": modification of work by "Pascal"/Flickr; credit "waterfall": modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter, but also with the structure of matter. Even a spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. On a molecular level, the bonds that hold the atoms of molecules together exist in a particular structure that has potential energy. Remember that anabolic

cellular pathways require energy to synthesize complex molecules from simpler ones and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called chemical energy. Chemical energy is responsible for providing living cells with energy from food. The release of energy occurs when the molecular bonds within food molecules are broken.

Note:
Concept in Action

Visit the site and select “Pendulum” from the “Work and Energy” menu to see the shifting kinetic and potential energy of a pendulum in motion.

Free and Activation Energy

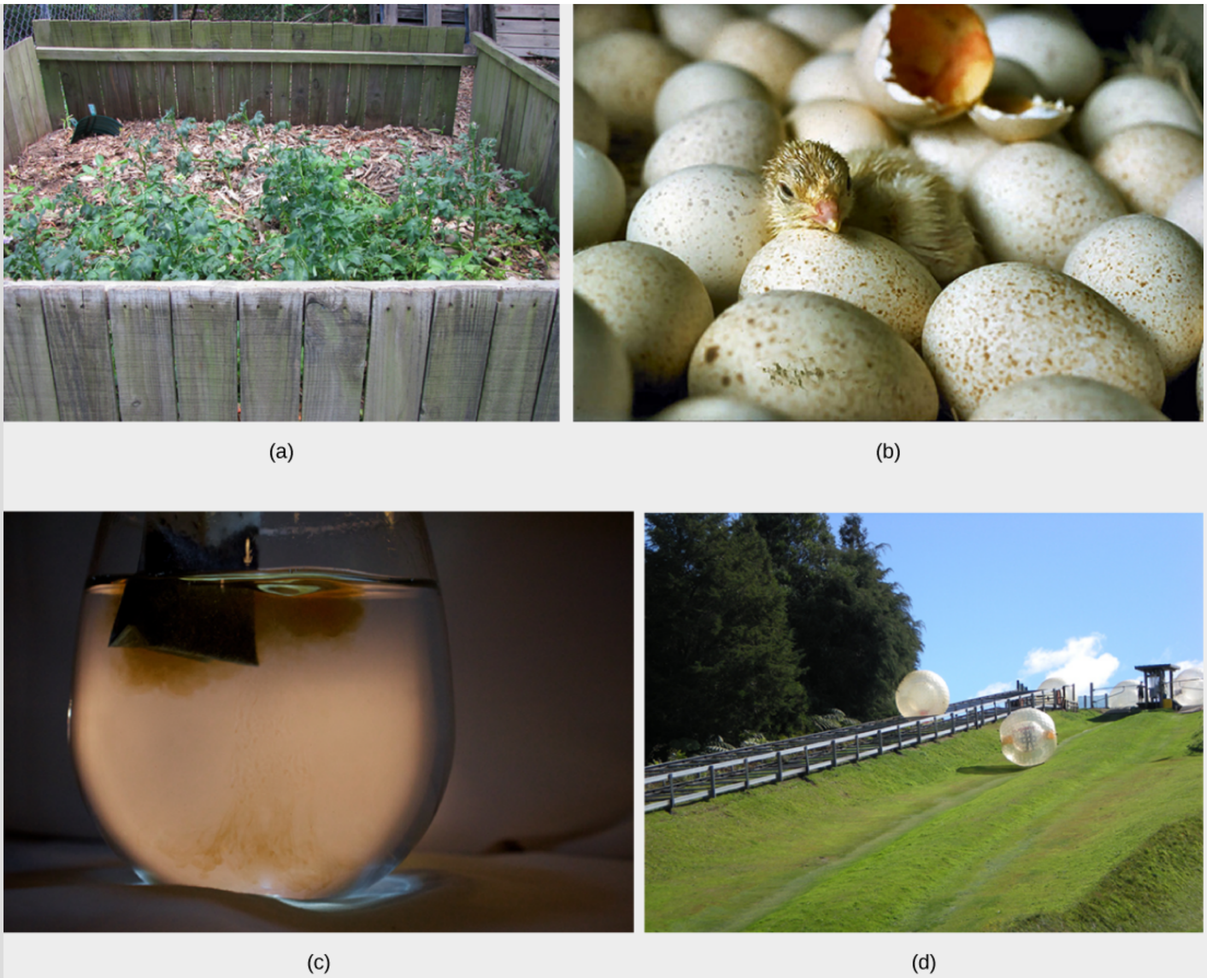
After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is the following: How is the energy associated with these chemical reactions quantified and expressed? How can the energy released from one reaction be compared to that of another reaction? A measurement of free energy is used to quantify these energy transfers. Recall that according to the second law of thermodynamics, all energy transfers involve the loss of some amount of energy in an unusable form such as heat. Free energy specifically refers to

the energy associated with a chemical reaction that is available after the losses are accounted for. In other words, free energy is usable energy, or energy that is available to do work.

If energy is released during a chemical reaction, then the change in free energy, signified as ΔG (delta G) will be a negative number. A negative change in free energy also means that the products of the reaction have less free energy than the reactants, because they release some free energy during the reaction. Reactions that have a negative change in free energy and consequently release free energy are called **exergonic reactions**. Think: *ex*ergonic means energy is *ex*iting the system. These reactions are also referred to as spontaneous reactions, and their products have less stored energy than the reactants. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction occurring immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction absorbs energy rather than releases energy on balance, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic reactions** and they are non-spontaneous. An endergonic reaction will not take place on its own without the addition of free energy.

Note:
Art Connection



Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by Cory Zanker; credit d: modification of work by Harry Malsch)

Look at each of the processes shown and decide if it is endergonic or exergonic.

There is another important concept that must be considered regarding endergonic and exergonic reactions. Exergonic reactions require a small amount of energy input to get going, before they can proceed with their

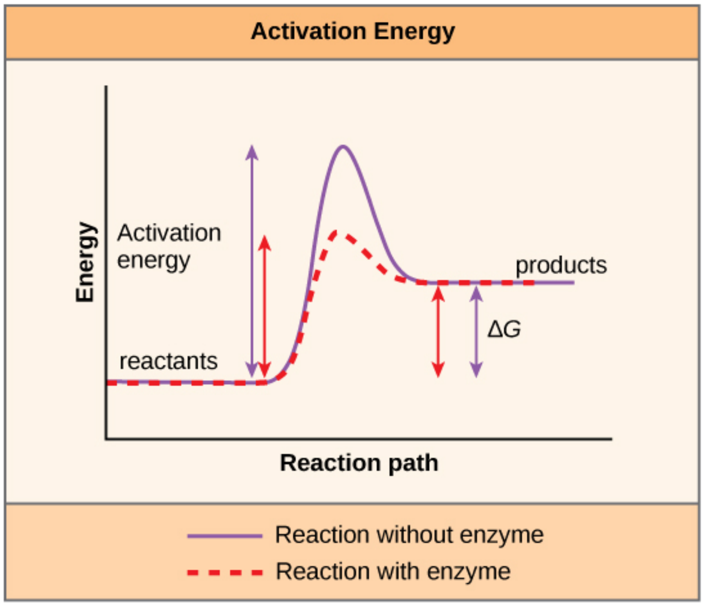
energy-releasing steps. These reactions have a net release of energy, but still require some energy input in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the **activation energy**.

Note:
Concept in Action

Watch an animation of the move from free energy to transition state of the reaction.

Enzymes

A substance that helps a chemical reaction to occur is called a catalyst, and the molecules that catalyze biochemical reactions are called **enzymes**. Most enzymes are proteins and perform the critical task of lowering the activation energies of chemical reactions inside the cell. Most of the reactions critical to a living cell happen too slowly at normal temperatures to be of any use to the cell. Without enzymes to speed up these reactions, life could not persist. Enzymes do this by binding to the reactant molecules and holding them in such a way as to make the chemical bond-breaking and -forming processes take place more easily. It is important to remember that enzymes do not change whether a reaction is exergonic (spontaneous) or endergonic. This is because they do not change the free energy of the reactants or products. They only reduce the activation energy required for the reaction to go forward ([link]). In addition, an enzyme itself is unchanged by the reaction it catalyzes. Once one reaction has been catalyzed, the enzyme is able to participate in other reactions.



Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

The chemical reactants to which an enzyme binds are called the enzyme's **substrates**. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a single reactant substrate is broken down into multiple products. In others, two substrates may come together to create one larger molecule. Two reactants might also enter a reaction and both become modified, but they leave the reaction as two products. The location within the enzyme where the substrate binds is called the enzyme's **active site**. The active site is where the "action" happens. Since enzymes are proteins, there is a unique combination of amino acid side chains within the active site. Each side chain is characterized by different properties. They can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of side chains creates a very specific chemical environment within the active site. This specific environment is suited to bind to one specific chemical substrate (or substrates).

Active sites are subject to influences of the local environment. Increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, temperatures outside of an optimal range reduce the rate at which an enzyme catalyzes a reaction. Hot temperatures will eventually cause enzymes to denature, an irreversible change in the three-dimensional shape and therefore the function of the enzyme. Enzymes are also suited to function best within a certain pH and salt concentration range, and, as with temperature, extreme pH, and salt concentrations can cause enzymes to denature.

For many years, scientists thought that enzyme-substrate binding took place in a simple “lock and key” fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a model called induced fit ([link]). The induced-fit model expands on the lock-and-key model by describing a more dynamic binding between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme’s structure that forms an ideal binding arrangement between enzyme and substrate.

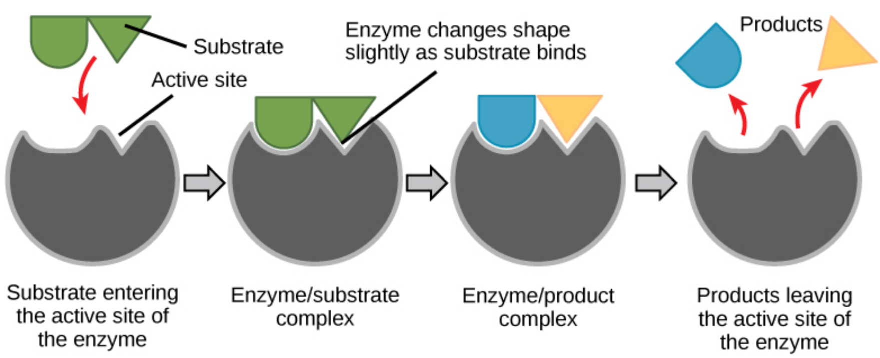
Note:
Concept in Action

View an animation of induced fit.

When an enzyme binds its substrate, an enzyme-substrate complex is formed. This complex lowers the activation energy of the reaction and promotes its rapid progression in one of multiple possible ways. On a basic level, enzymes promote chemical reactions that involve more than one

substrate by bringing the substrates together in an optimal orientation for reaction. Another way in which enzymes promote the reaction of their substrates is by creating an optimal environment within the active site for the reaction to occur. The chemical properties that emerge from the particular arrangement of amino acid R groups within an active site create the perfect environment for an enzyme's specific substrates to react.

The enzyme-substrate complex can also lower activation energy by compromising the bond structure so that it is easier to break. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. In these cases, it is important to remember that the enzyme will always return to its original state by the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme has catalyzed a reaction, it releases its product(s) and can catalyze a new reaction.



The induced-fit model is an adjustment to the lock-and-key model and explains how enzymes and substrates undergo dynamic modifications during the transition state to increase the affinity of the substrate for the active site.

It would seem ideal to have a scenario in which all of an organism's enzymes existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. However, a variety of

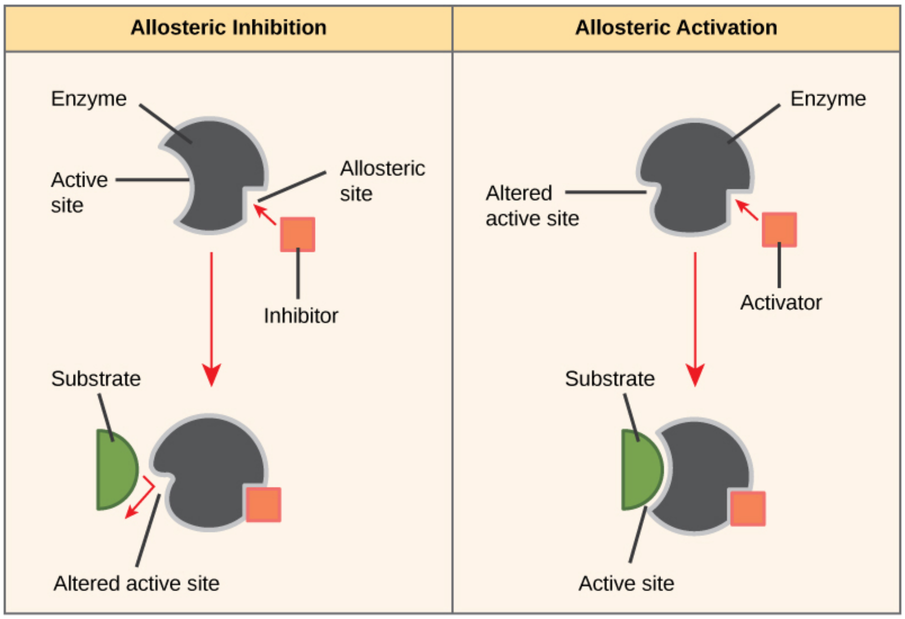
mechanisms ensures that this does not happen. Cellular needs and conditions constantly vary from cell to cell, and change within individual cells over time. The required enzymes of stomach cells differ from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive organ cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so must the amounts and functionality of different enzymes.

Since the rates of biochemical reactions are controlled by activation energy, and enzymes lower and determine activation energies for chemical reactions, the relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at what rates. This determination is tightly controlled in cells. In certain cellular environments, enzyme activity is partly controlled by environmental factors like pH, temperature, salt concentration, and, in some cases, cofactors or coenzymes.

Enzymes can also be regulated in ways that either promote or reduce enzyme activity. There are many kinds of molecules that inhibit or promote enzyme function, and various mechanisms by which they do so. In some cases of enzyme inhibition, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for binding to the active site.

On the other hand, in **noncompetitive inhibition**, an inhibitor molecule binds to the enzyme in a location other than the active site, called an allosteric site, but still manages to block substrate binding to the active site. Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for its substrate. This type of inhibition is called **allosteric inhibition** ([link]). Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to a region on an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates

with less efficiency. There are allosteric activators as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s) ([link]).



Allosteric inhibition works by indirectly inducing a conformational change to the active site such that the substrate no longer fits. In contrast, in allosteric activation, the activator molecule modifies the shape of the active site to allow a better fit of the substrate.

Note:
Careers in Action
Pharmaceutical Drug Developer

Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated are key principles behind the development of many of the pharmaceutical drugs on the market today. Biologists working in this field collaborate with other scientists to design drugs ([link]).

Consider statins for example—statins is the name given to one class of drugs that can reduce cholesterol levels. These compounds are inhibitors of the enzyme HMG-CoA reductase, which is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the level of cholesterol synthesized in the body can be reduced. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is used to provide relief from fever and inflammation (pain), its mechanism of action is still not completely understood.

How are drugs discovered? One of the biggest challenges in drug discovery is identifying a drug target. A drug target is a molecule that is literally the target of the drug. In the case of statins, HMG-CoA reductase is the drug target. Drug targets are identified through painstaking research in the laboratory. Identifying the target alone is not enough; scientists also need to know how the target acts inside the cell and which reactions go awry in the case of disease. Once the target and the pathway are identified, then the actual process of drug design begins. In this stage, chemists and

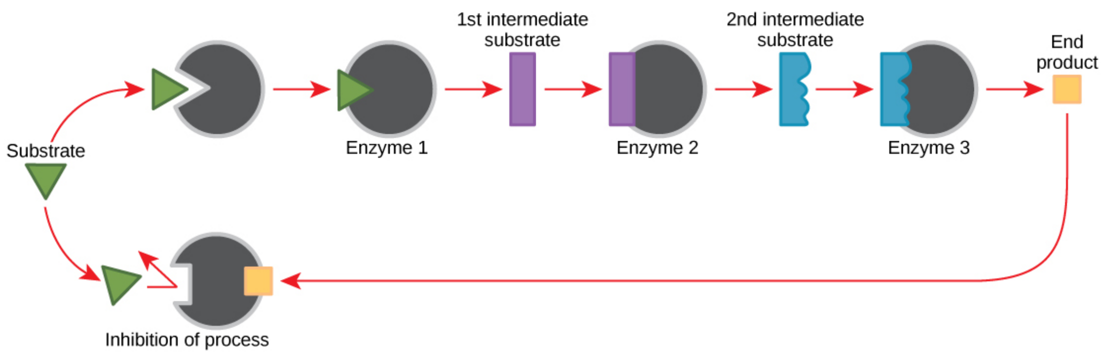
biologists work together to design and synthesize molecules that can block or activate a particular reaction. However, this is only the beginning: If and when a drug prototype is successful in performing its function, then it is subjected to many tests from in vitro experiments to clinical trials before it can get approval from the U.S. Food and Drug Administration to be on the market.

Many enzymes do not work optimally, or even at all, unless bound to other specific non-protein helper molecules. They may bond either temporarily through ionic or hydrogen bonds, or permanently through stronger covalent bonds. Binding to these molecules promotes optimal shape and function of their respective enzymes. Two examples of these types of helper molecules are cofactors and coenzymes. Cofactors are inorganic ions such as ions of iron and magnesium. Coenzymes are organic helper molecules, those with a basic atomic structure made up of carbon and hydrogen. Like enzymes, these molecules participate in reactions without being changed themselves and are ultimately recycled and reused. Vitamins are the source of coenzymes. Some vitamins are the precursors of coenzymes and others act directly as coenzymes. Vitamin C is a direct coenzyme for multiple enzymes that take part in building the important connective tissue, collagen. Therefore, enzyme function is, in part, regulated by the abundance of various cofactors and coenzymes, which may be supplied by an organism's diet or, in some cases, produced by the organism.

Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. The major question remains, however: What are these molecules and where do they come from? Some are cofactors and coenzymes, as you have learned. What other molecules in the cell provide enzymatic regulation such as allosteric modulation, and competitive and non-competitive inhibition? Perhaps the most relevant sources of regulatory molecules, with respect to enzymatic cellular metabolism, are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use

the products of their own reactions for feedback inhibition of enzyme activity. **Feedback inhibition** involves the use of a reaction product to regulate its own further production ([link]). The cell responds to an abundance of the products by slowing down production during anabolic or catabolic reactions. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms described above.



Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream process, is an important regulatory mechanism in cells.

The production of both amino acids and nucleotides is controlled through feedback inhibition. Additionally, ATP is an allosteric regulator of some of the enzymes involved in the catabolic breakdown of sugar, the process that creates ATP. In this way, when ATP is in abundant supply, the cell can prevent the production of ATP. On the other hand, ADP serves as a positive allosteric regulator (an allosteric activator) for some of the same enzymes that are inhibited by ATP. Thus, when relative levels of ADP are high compared to ATP, the cell is triggered to produce more ATP through sugar catabolism.

Section Summary

Cells perform the functions of life through various chemical reactions. A cell's metabolism refers to the combination of chemical reactions that take place within it. Catabolic reactions break down complex chemicals into simpler ones and are associated with energy release. Anabolic processes build complex molecules out of simpler ones and require energy.

In studying energy, the term system refers to the matter and environment involved in energy transfers. Entropy is a measure of the disorder of a system. The physical laws that describe the transfer of energy are the laws of thermodynamics. The first law states that the total amount of energy in the universe is constant. The second law of thermodynamics states that every energy transfer involves some loss of energy in an unusable form, such as heat energy. Energy comes in different forms: kinetic, potential, and free. The change in free energy of a reaction can be negative (releases energy, exergonic) or positive (consumes energy, endergonic). All reactions require an initial input of energy to proceed, called the activation energy.

Enzymes are chemical catalysts that speed up chemical reactions by lowering their activation energy. Enzymes have an active site with a unique chemical environment that fits particular chemical reactants for that enzyme, called substrates. Enzymes and substrates are thought to bind according to an induced-fit model. Enzyme action is regulated to conserve resources and respond optimally to the environment.

Art Connections

Exercise:

Problem:

[link] Look at each of the processes shown and decide if it is endergonic or exergonic.

Solution:

[link] A compost pile decomposing is an exergonic process. A baby developing from a fertilized egg is an endergonic process. Tea

dissolving into water is an exergonic process. A ball rolling downhill is an exergonic process.

Review Questions

Exercise:

Problem:

Which of the following is not an example of an energy transformation?

a. Heating up dinner in a microwave
b. Solar panels at work
c. Formation of static electricity
d. None of the above

Solution:

D

Exercise:

Problem: Which of the following is not true about enzymes?

a. They are consumed by the reactions they catalyze.
b. They are usually made of amino acids.
c. They lower the activation energy of chemical reactions.
d. Each one is specific to the particular substrate(s) to which it binds.

Solution:

A

Free Response

Exercise:

Problem:

Does physical exercise to increase muscle mass involve anabolic and/or catabolic processes? Give evidence for your answer.

Solution:

Physical exercise involves both anabolic and catabolic processes. Body cells break down sugars to provide ATP to do the work necessary for exercise, such as muscle contractions. This is catabolism. Muscle cells also must repair muscle tissue damaged by exercise by building new muscle. This is anabolism.

Exercise:

Problem:

Explain in your own terms the difference between a spontaneous reaction and one that occurs instantaneously, and what causes this difference.

Solution:

A spontaneous reaction is one that has a negative ΔG and thus releases energy. However, a spontaneous reaction need not occur quickly or suddenly like an instantaneous reaction. It may occur over long periods of time due to a large energy of activation, which prevents the reaction from occurring quickly.

Exercise:

Problem:

With regard to enzymes, why are vitamins and minerals necessary for good health? Give examples.

Solution:

Most vitamins and minerals act as cofactors and coenzymes for enzyme action. Many enzymes require the binding of certain cofactors or coenzymes to be able to catalyze their reactions. Since enzymes catalyze many important reactions, it is critical to obtain sufficient vitamins and minerals from diet and supplements. Vitamin C (ascorbic acid) is a coenzyme necessary for the action of enzymes that build collagen.

Glossary

activation energy
: the amount of initial energy necessary for reactions to occur

active site
: a specific region on the enzyme where the substrate binds

allosteric inhibition
: the mechanism for inhibiting enzyme action in which a regulatory molecule binds to a second site (not the active site) and initiates a conformation change in the active site, preventing binding with the substrate

anabolic
: describes the pathway that requires a net energy input to synthesize complex molecules from simpler ones

bioenergetics
: the concept of energy flow through living systems

catabolic
: describes the pathway in which complex molecules are broken down into simpler ones, yielding energy as an additional product of the reaction

competitive inhibition
: a general mechanism of enzyme activity regulation in which a molecule other than the enzyme’s substrate is able to bind the active

site and prevent the substrate itself from binding, thus inhibiting the overall rate of reaction for the enzyme

endergonic
: describes a chemical reaction that results in products that store more chemical potential energy than the reactants

enzyme
: a molecule that catalyzes a biochemical reaction

exergonic
: describes a chemical reaction that results in products with less chemical potential energy than the reactants, plus the release of free energy

feedback inhibition
: a mechanism of enzyme activity regulation in which the product of a reaction or the final product of a series of sequential reactions inhibits an enzyme for an earlier step in the reaction series

heat energy
: the energy transferred from one system to another that is not work

kinetic energy
: the type of energy associated with objects in motion

metabolism
: all the chemical reactions that take place inside cells, including those that use energy and those that release energy

noncompetitive inhibition
: a general mechanism of enzyme activity regulation in which a regulatory molecule binds to a site other than the active site and prevents the active site from binding the substrate; thus, the inhibitor molecule does not compete with the substrate for the active site; allosteric inhibition is a form of noncompetitive inhibition

potential energy

the type of energy that refers to the potential to do work

substrate
: a molecule on which the enzyme acts

thermodynamics
: the science of the relationships between heat, energy, and work

Glycolysis

By the end of this section, you will be able to:

- Explain how ATP is used by the cell as an energy source
- Describe the overall result in terms of molecules produced of the breakdown of glucose by glycolysis

Even exergonic, energy-releasing reactions require a small amount of activation energy to proceed. However, consider endergonic reactions, which require much more energy input because their products have more free energy than their reactants. Within the cell, where does energy to power such reactions come from? The answer lies with an energy-supplying molecule called adenosine triphosphate, or **ATP**. ATP is a small, relatively simple molecule, but within its bonds contains the potential for a quick burst of energy that can be harnessed to perform cellular work. This molecule can be thought of as the primary energy currency of cells in the same way that money is the currency that people exchange for things they need. ATP is used to power the majority of energy-requiring cellular reactions.

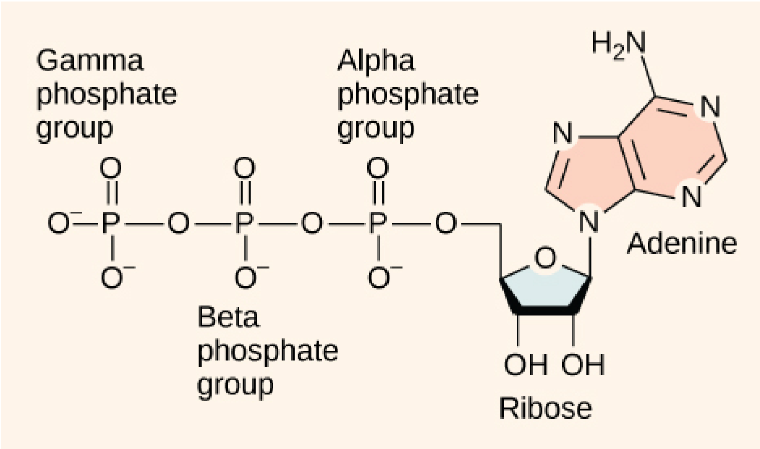
ATP in Living Systems

A living cell cannot store significant amounts of free energy. Excess free energy would result in an increase of heat in the cell, which would denature enzymes and other proteins, and thus destroy the cell. Rather, a cell must be able to store energy safely and release it for use only as needed. Living cells accomplish this using ATP, which can be used to fill any energy need of the cell. How? It functions as a rechargeable battery.

When ATP is broken down, usually by the removal of its terminal phosphate group, energy is released. This energy is used to do work by the cell, usually by the binding of the released phosphate to another molecule, thus activating it. For example, in the mechanical work of muscle contraction, ATP supplies energy to move the contractile muscle proteins.

ATP Structure and Function

At the heart of ATP is a molecule of adenosine monophosphate (AMP), which is composed of an adenine molecule bonded to both a ribose molecule and a single phosphate group ([link]). Ribose is a five-carbon sugar found in RNA and AMP is one of the nucleotides in RNA. The addition of a second phosphate group to this core molecule results in adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).



The structure of ATP shows the basic components of a two-ring adenine, five-carbon ribose, and three phosphate groups.

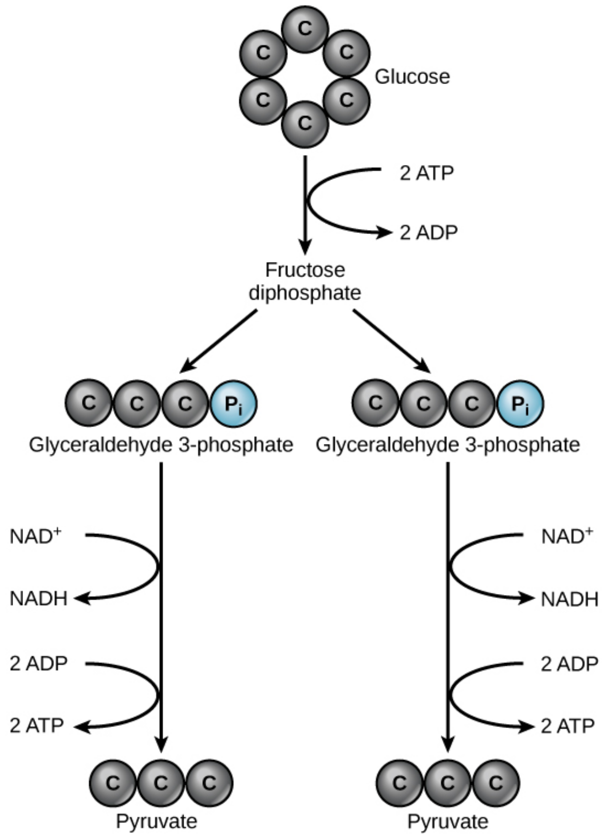
The addition of a phosphate group to a molecule requires a high amount of energy and results in a high-energy bond. Phosphate groups are negatively charged and thus repel one another when they are arranged in series, as they are in ADP and ATP. This repulsion makes the ADP and ATP molecules inherently unstable. The release of one or two phosphate groups from ATP, a process called hydrolysis, releases energy.

Glycolysis

You have read that nearly all of the energy used by living things comes to them in the bonds of the sugar, glucose. **Glycolysis** is the first step in the breakdown of glucose to extract energy for cell metabolism. Many living organisms carry out glycolysis as part of their metabolism. Glycolysis takes place in the cytoplasm of most prokaryotic and all eukaryotic cells.

Glycolysis begins with the six-carbon, ring-shaped structure of a single glucose molecule and ends with two molecules of a three-carbon sugar called pyruvate. Glycolysis consists of two distinct phases. In the first part of the glycolysis pathway, energy is used to make adjustments so that the six-carbon sugar molecule can be split evenly into two three-carbon pyruvate molecules. In the second part of glycolysis, ATP and nicotinamide-adenine dinucleotide (NADH) are produced ([link]).

If the cell cannot catabolize the pyruvate molecules further, it will harvest only two ATP molecules from one molecule of glucose. For example, mature mammalian red blood cells are only capable of glycolysis, which is their sole source of ATP. If glycolysis is interrupted, these cells would eventually die.



In glycolysis, a glucose molecule is converted into two pyruvate molecules.

Section Summary

ATP functions as the energy currency for cells. It allows cells to store energy briefly and transport it within itself to support endergonic chemical reactions. The structure of ATP is that of an RNA nucleotide with three phosphate groups attached. As ATP is used for energy, a phosphate group is detached, and ADP is produced. Energy derived from glucose catabolism is used to recharge ADP into ATP.

Glycolysis is the first pathway used in the breakdown of glucose to extract energy. Because it is used by nearly all organisms on earth, it must have

evolved early in the history of life. Glycolysis consists of two parts: The first part prepares the six-carbon ring of glucose for separation into two three-carbon sugars. Energy from ATP is invested into the molecule during this step to energize the separation. The second half of glycolysis extracts ATP and high-energy electrons from hydrogen atoms and attaches them to NAD^+. Two ATP molecules are invested in the first half and four ATP molecules are formed during the second half. This produces a net gain of two ATP molecules per molecule of glucose for the cell.

Multiple Choice

Exercise:

Problem:

Energy is stored long-term in the bonds of ______ and used short-term to perform work from a(n) ______ molecule.

a. ATP : glucose
b. an anabolic molecule : catabolic molecule
c. glucose : ATP
d. a catabolic molecule : anabolic molecule

Solution:

C

Exercise:

Problem: The energy currency used by cells is ______.

a. ATP
b. ADP
c. AMP
d. adenosine

Solution:

A

Exercise:

Problem:

The glucose that enters the glycolysis pathway is split into two molecules of __________.

a. ATP
b. phosphate
c. NADH
d. pyruvate

Solution:

D

Free Response

Exercise:

Problem:

Both prokaryotic and eukaryotic organisms carry out some form of glycolysis. How does that fact support or not support the assertion that glycolysis is one of the oldest metabolic pathways?

Solution:

If glycolysis evolved relatively late, it likely would not be as universal in organisms as it is. It probably evolved in very primitive organisms and persisted, with the addition of other pathways of carbohydrate metabolism that evolved later.

Glossary

ATP
: (also, adenosine triphosphate) the cell’s energy currency

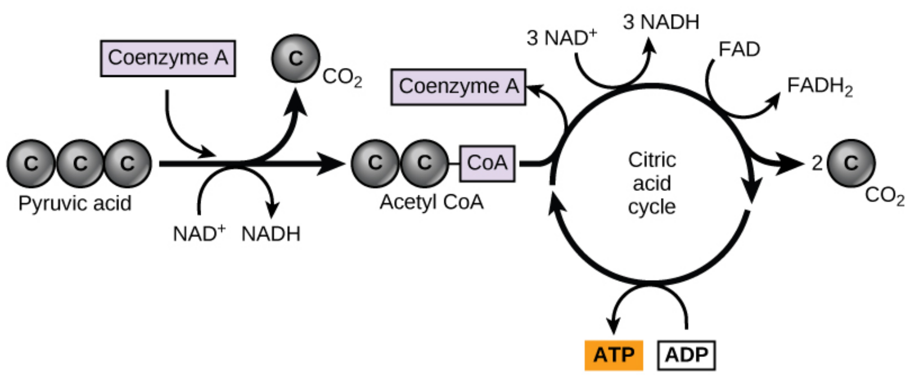
glycolysis
: the process of breaking glucose into two three-carbon molecules with the production of ATP and NADH

Citric Acid Cycle and Oxidative Phosphorylation
By the end of this section, you will be able to:

- Describe the location of the citric acid cycle and oxidative phosphorylation in the cell
- Describe the overall outcome of the citric acid cycle and oxidative phosphorylation in terms of the products of each
- Describe the relationships of glycolysis, the citric acid cycle, and oxidative phosphorylation in terms of their inputs and outputs.

The Citric Acid Cycle

In eukaryotic cells, the pyruvate molecules produced at the end of glycolysis are transported into mitochondria, which are sites of cellular respiration. If oxygen is available, aerobic respiration will go forward. In mitochondria, pyruvate will be transformed into a two-carbon acetyl group (by removing a molecule of carbon dioxide) that will be picked up by a carrier compound called coenzyme A (CoA), which is made from vitamin B_5. The resulting compound is called **acetyl CoA**. ([link]). Acetyl CoA can be used in a variety of ways by the cell, but its major function is to deliver the acetyl group derived from pyruvate to the next pathway in glucose catabolism.



Pyruvate is converted into acetyl-CoA before entering the citric acid cycle.

Like the conversion of pyruvate to acetyl CoA, the **citric acid cycle** in eukaryotic cells takes place in the matrix of the mitochondria. Unlike glycolysis, the citric acid cycle is a closed loop: The last part of the pathway regenerates the compound used in the first step. The eight steps of the cycle are a series of chemical reactions that produces two carbon dioxide molecules, one ATP molecule (or an equivalent), and reduced forms (NADH and $FADH_2$) of NAD^+ and FAD^+, important coenzymes in the cell. Part of this is considered an aerobic pathway (oxygen-requiring) because the NADH and $FADH_2$ produced must transfer their electrons to the next pathway in the system, which will use oxygen. If oxygen is not present, this transfer does not occur.

Two carbon atoms come into the citric acid cycle from each acetyl group. Two carbon dioxide molecules are released on each turn of the cycle; however, these do not contain the same carbon atoms contributed by the acetyl group on that turn of the pathway. The two acetyl-carbon atoms will eventually be released on later turns of the cycle; in this way, all six carbon atoms from the original glucose molecule will be eventually released as carbon dioxide. It takes two turns of the cycle to process the equivalent of one glucose molecule. Each turn of the cycle forms three high-energy NADH molecules and one high-energy $FADH_2$ molecule. These high-energy carriers will connect with the last portion of aerobic respiration to produce ATP molecules. One ATP (or an equivalent) is also made in each cycle. Several of the intermediate compounds in the citric acid cycle can be used in synthesizing non-essential amino acids; therefore, the cycle is both anabolic and catabolic.

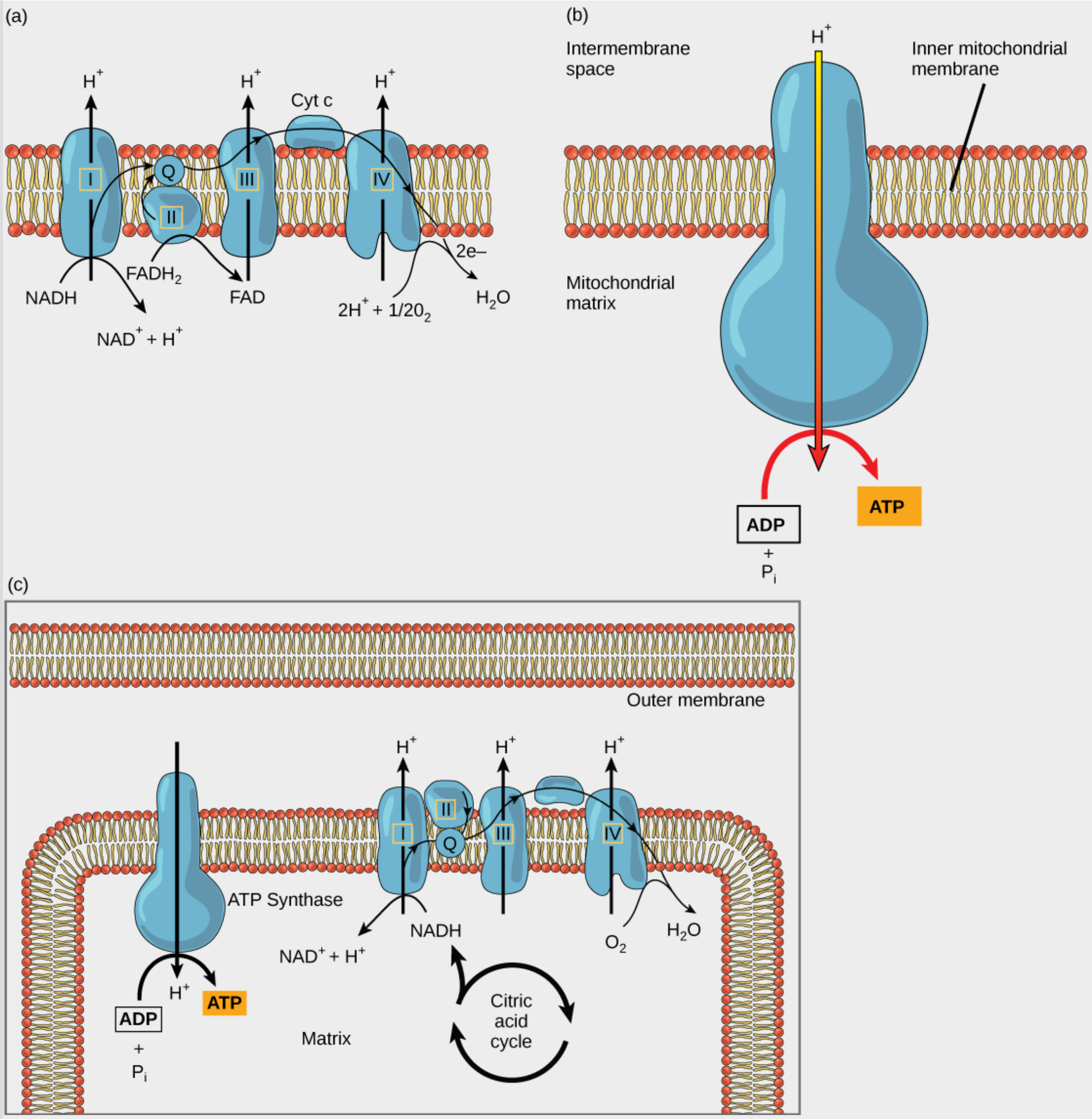
Oxidative Phosphorylation

You have just read about two pathways in glucose catabolism—glycolysis and the citric acid cycle—that generate ATP. Most of the ATP generated during the aerobic catabolism of glucose, however, is not generated directly from these pathways. Rather, it derives from a process that begins with passing electrons through a series of chemical reactions to a final electron acceptor, oxygen. These reactions take place in specialized protein

complexes located in the inner membrane of the mitochondria of eukaryotic organisms and on the inner part of the cell membrane of prokaryotic organisms. The energy of the electrons is harvested and used to generate a electrochemical gradient across the inner mitochondrial membrane. The potential energy of this gradient is used to generate ATP. The entirety of this process is called **oxidative phosphorylation**.

The electron transport chain ([link]**a**) is the last component of aerobic respiration and is the only part of metabolism that uses atmospheric oxygen. Oxygen continuously diffuses into plants for this purpose. In animals, oxygen enters the body through the respiratory system. Electron transport is a series of chemical reactions that resembles a bucket brigade in that electrons are passed rapidly from one component to the next, to the endpoint of the chain where oxygen is the final electron acceptor and water is produced. There are four complexes composed of proteins, labeled I through IV in [link]**c**, and the aggregation of these four complexes, together with associated mobile, accessory electron carriers, is called the **electron transport chain**. The electron transport chain is present in multiple copies in the inner mitochondrial membrane of eukaryotes and in the plasma membrane of prokaryotes. In each transfer of an electron through the electron transport chain, the electron loses energy, but with some transfers, the energy is stored as potential energy by using it to pump hydrogen ions across the inner mitochondrial membrane into the intermembrane space, creating an electrochemical gradient.

Note:
Art Connection



(a) The electron transport chain is a set of molecules that supports a series of oxidation-reduction reactions. (b) ATP synthase is a complex, molecular machine that uses an H^+ gradient to regenerate ATP from ADP. (c) Chemiosmosis relies on the potential energy provided by the H^+ gradient across the membrane.

Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of

the intermembrane space to increase or decrease? What affect would cyanide have on ATP synthesis?

Electrons from NADH and $FADH_2$ are passed to protein complexes in the electron transport chain. As they are passed from one complex to another (there are a total of four), the electrons lose energy, and some of that energy is used to pump hydrogen ions from the mitochondrial matrix into the intermembrane space. In the fourth protein complex, the electrons are accepted by oxygen, the terminal acceptor. The oxygen with its extra electrons then combines with two hydrogen ions, further enhancing the electrochemical gradient, to form water. If there were no oxygen present in the mitochondrion, the electrons could not be removed from the system, and the entire electron transport chain would back up and stop. The mitochondria would be unable to generate new ATP in this way, and the cell would ultimately die from lack of energy. This is the reason we must breathe to draw in new oxygen.

In the electron transport chain, the free energy from the series of reactions just described is used to pump hydrogen ions across the membrane. The uneven distribution of H^+ ions across the membrane establishes an electrochemical gradient, owing to the H^+ ions' positive charge and their higher concentration on one side of the membrane.

Hydrogen ions diffuse through the inner membrane through an integral membrane protein called **ATP synthase** ([link]**b**). This complex protein acts as a tiny generator, turned by the force of the hydrogen ions diffusing through it, down their electrochemical gradient from the intermembrane space, where there are many mutually repelling hydrogen ions to the matrix, where there are few. The turning of the parts of this molecular machine regenerate ATP from ADP. This flow of hydrogen ions across the membrane through ATP synthase is called **chemiosmosis**.

Chemiosmosis ([link]**c**) is used to generate 90 percent of the ATP made during aerobic glucose catabolism. The result of the reactions is the production of ATP from the energy of the electrons removed from hydrogen atoms. These atoms were originally part of a glucose molecule. At the end

of the electron transport system, the electrons are used to reduce an oxygen molecule to oxygen ions. The extra electrons on the oxygen ions attract hydrogen ions (protons) from the surrounding medium, and water is formed. The electron transport chain and the production of ATP through chemiosmosis are collectively called oxidative phosphorylation.

ATP Yield

The number of ATP molecules generated from the catabolism of glucose varies. For example, the number of hydrogen ions that the electron transport chain complexes can pump through the membrane varies between species. Another source of variance stems from the shuttle of electrons across the mitochondrial membrane. The NADH generated from glycolysis cannot easily enter mitochondria. Thus, electrons are picked up on the inside of the mitochondria by either NAD^+ or FAD^+. Fewer ATP molecules are generated when FAD^+ acts as a carrier. NAD^+ is used as the electron transporter in the liver and FAD^+ in the brain, so ATP yield depends on the tissue being considered.

Another factor that affects the yield of ATP molecules generated from glucose is that intermediate compounds in these pathways are used for other purposes. Glucose catabolism connects with the pathways that build or break down all other biochemical compounds in cells, and the result is somewhat messier than the ideal situations described thus far. For example, sugars other than glucose are fed into the glycolytic pathway for energy extraction. Other molecules that would otherwise be used to harvest energy in glycolysis or the citric acid cycle may be removed to form nucleic acids, amino acids, lipids, or other compounds. Overall, in living systems, these pathways of glucose catabolism extract about 34 percent of the energy contained in glucose.

Note:
Careers in Action
Mitochondrial Disease Physician

What happens when the critical reactions of cellular respiration do not proceed correctly? Mitochondrial diseases are genetic disorders of metabolism. Mitochondrial disorders can arise from mutations in nuclear or mitochondrial DNA, and they result in the production of less energy than is normal in body cells. Symptoms of mitochondrial diseases can include muscle weakness, lack of coordination, stroke-like episodes, and loss of vision and hearing. Most affected people are diagnosed in childhood, although there are some adult-onset diseases. Identifying and treating mitochondrial disorders is a specialized medical field. The educational preparation for this profession requires a college education, followed by medical school with a specialization in medical genetics. Medical geneticists can be board certified by the American Board of Medical Genetics and go on to become associated with professional organizations devoted to the study of mitochondrial disease, such as the Mitochondrial Medicine Society and the Society for Inherited Metabolic Disease.

Section Summary

The citric acid cycle is a series of chemical reactions that removes high-energy electrons and uses them in the electron transport chain to generate ATP. One molecule of ATP (or an equivalent) is produced per each turn of the cycle.

The electron transport chain is the portion of aerobic respiration that uses free oxygen as the final electron acceptor for electrons removed from the intermediate compounds in glucose catabolism. The electrons are passed through a series of chemical reactions, with a small amount of free energy used at three points to transport hydrogen ions across the membrane. This contributes to the gradient used in chemiosmosis. As the electrons are passed from NADH or $FADH_2$ down the electron transport chain, they lose energy. The products of the electron transport chain are water and ATP. A number of intermediate compounds can be diverted into the anabolism of other biochemical molecules, such as nucleic acids, non-essential amino

acids, sugars, and lipids. These same molecules, except nucleic acids, can serve as energy sources for the glucose pathway.

Art Connections

Exercise:

Problem:

[link] Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What affect would cyanide have on ATP synthesis?

Solution:

[link] After cyanide poisoning, the electron transport chain can no longer pump electrons into the intermembrane space. The pH of the intermembrane space would increase, and ATP synthesis would stop.

Multiple Choice

Exercise:

Problem: What do the electrons added to NAD^+ do?

a. They become part of a fermentation pathway.
b. They go to another pathway for ATP production.
c. They energize the entry of the acetyl group into the citric acid cycle.
d. They are converted into NADP.

Solution:

B

Exercise:

Problem: Chemiosmosis involves

a. the movement of electrons across the cell membrane
b. the movement of hydrogen atoms across a mitochondrial membrane
c. the movement of hydrogen ions across a mitochondrial membrane
d. the movement of glucose through the cell membrane

Solution:

C

Free Response

Exercise:

Problem:

We inhale oxygen when we breathe and exhale carbon dioxide. What is the oxygen used for and where does the carbon dioxide come from?

Solution:

The oxygen we inhale is the final electron acceptor in the electron transport chain and allows aerobic respiration to proceed, which is the most efficient pathway for harvesting energy in the form of ATP from food molecules. The carbon dioxide we breathe out is formed during the citric acid cycle when the bonds in carbon compounds are broken.

Glossary

acetyl CoA

the combination of an acetyl group derived from pyruvic acid and coenzyme A which is made from pantothenic acid (a B-group vitamin)

ATP synthase
: a membrane-embedded protein complex that regenerates ATP from ADP with energy from protons diffusing through it

chemiosmosis
: the movement of hydrogen ions down their electrochemical gradient across a membrane through ATP synthase to generate ATP

citric acid cycle
: a series of enzyme-catalyzed chemical reactions of central importance in all living cells that harvests the energy in carbon-carbon bonds of sugar molecules to generate ATP; the citric acid cycle is an aerobic metabolic pathway because it requires oxygen in later reactions to proceed

electron transport chain
: a series of four large, multi-protein complexes embedded in the inner mitochondrial membrane that accepts electrons from donor compounds and harvests energy from a series of chemical reactions to generate a hydrogen ion gradient across the membrane

oxidative phosphorylation
: the production of ATP by the transfer of electrons down the electron transport chain to create a proton gradient that is used by ATP synthase to add phosphate groups to ADP molecules

Fermentation
By the end of this section, you will be able to:

- Discuss the fundamental difference between anaerobic cellular respiration and fermentation
- Describe the type of fermentation that readily occurs in animal cells and the conditions that initiate that fermentation

In aerobic respiration, the final electron acceptor is an oxygen molecule, O_2. If aerobic respiration occurs, then ATP will be produced using the energy of the high-energy electrons carried by NADH or $FADH_2$ to the electron transport chain. If aerobic respiration does not occur, NADH must be reoxidized to NAD^+ for reuse as an electron carrier for glycolysis to continue. How is this done? Some living systems use an organic molecule as the final electron acceptor. Processes that use an organic molecule to regenerate NAD^+ from NADH are collectively referred to as **fermentation**. In contrast, some living systems use an inorganic molecule (other than oxygen) as a final electron acceptor to regenerate NAD^+; both methods are anaerobic (do not require oxygen) to achieve NAD^+ regeneration and enable organisms to convert energy for their use in the absence of oxygen.

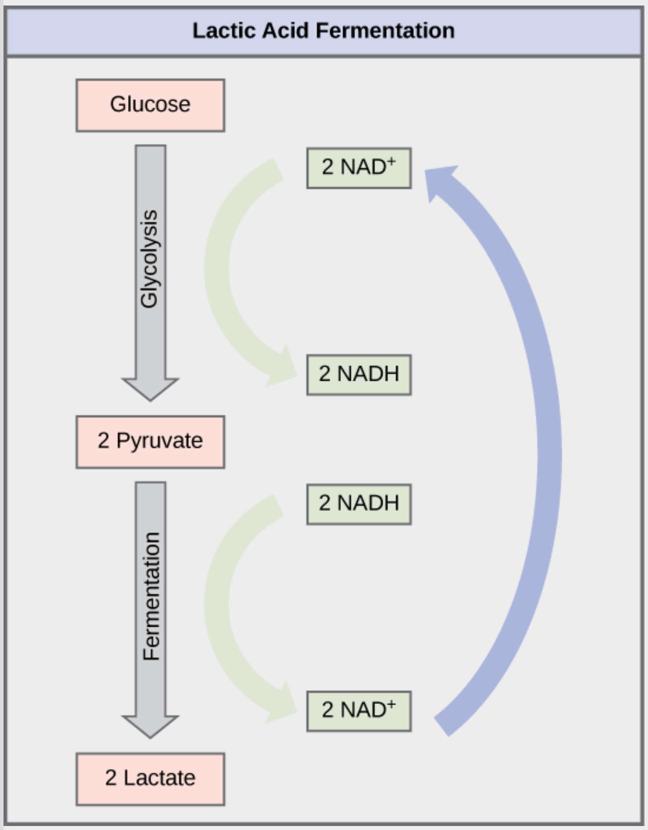
Lactic Acid Fermentation

The fermentation method used by animals and some bacteria like those in yogurt is lactic acid fermentation ([link]). This occurs routinely in mammalian red blood cells and in skeletal muscle that has insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). In muscles, lactic acid produced by fermentation must be removed by the blood circulation and brought to the liver for further metabolism. The chemical reaction of lactic acid fermentation is the following:
Equation:

$$\text{Pyruvic acid } + \text{NADH} \leftrightarrow \text{lactic acid} + \text{NAD}^{+}$$

The enzyme that catalyzes this reaction is lactate dehydrogenase. The reaction can proceed in either direction, but the left-to-right reaction is inhibited by acidic conditions. This lactic acid build-up causes muscle stiffness and fatigue. Once the lactic acid has been removed from the muscle and is circulated to the liver, it can be converted back to pyruvic acid and further catabolized for energy.

Note:
Art Connection



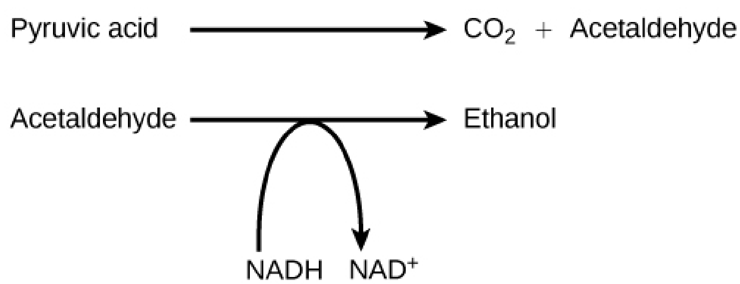
Lactic acid fermentation is common in muscles that have become exhausted by use.

Tremetol, a metabolic poison found in white snake root plant, prevents the metabolism of lactate. When cows eat this plant, Tremetol is concentrated

in the milk. Humans who consume the milk become ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

Alcohol Fermentation

Another familiar fermentation process is alcohol fermentation ([link]), which produces ethanol, an alcohol. The alcohol fermentation reaction is the following:



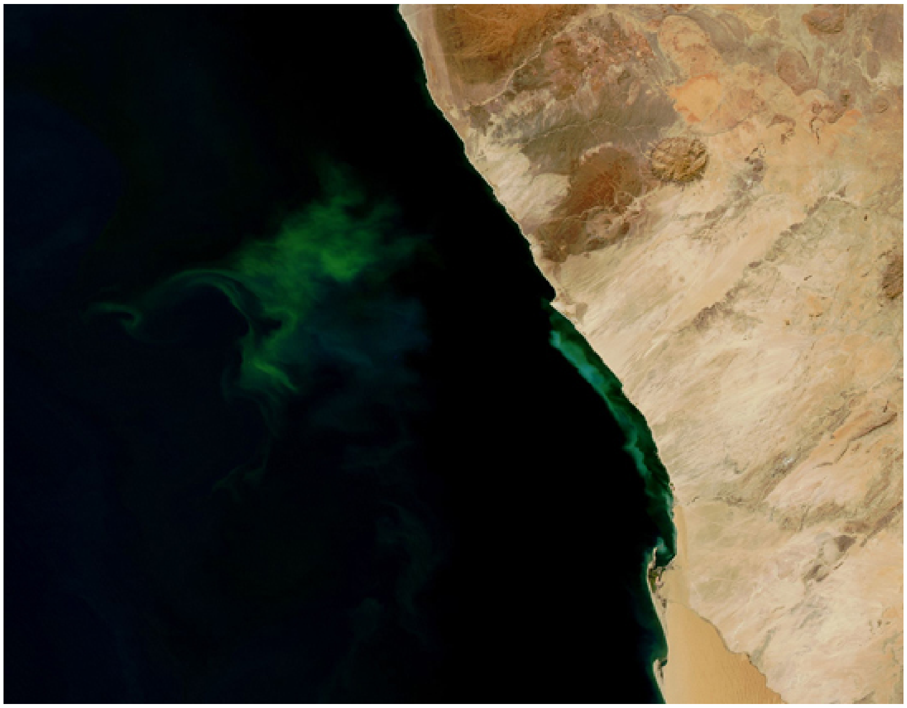
The reaction resulting in alcohol fermentation is shown.

In the first reaction, a carboxyl group is removed from pyruvic acid, releasing carbon dioxide as a gas. The loss of carbon dioxide reduces the molecule by one carbon atom, making acetaldehyde. The second reaction removes an electron from NADH, forming NAD^+ and producing ethanol from the acetaldehyde, which accepts the electron. The fermentation of pyruvic acid by yeast produces the ethanol found in alcoholic beverages ([link]). If the carbon dioxide produced by the reaction is not vented from the fermentation chamber, for example in beer and sparkling wines, it remains dissolved in the medium until the pressure is released. Ethanol above 12 percent is toxic to yeast, so natural levels of alcohol in wine occur at a maximum of 12 percent.

Fermentation of grape juice to make wine produces CO_2 as a byproduct. Fermentation tanks have valves so that pressure inside the tanks can be released.

Anaerobic Cellular Respiration

Certain prokaryotes, including some species of bacteria and Archaea, use anaerobic respiration. For example, the group of Archaea called methanogens reduces carbon dioxide to methane to oxidize NADH. These microorganisms are found in soil and in the digestive tracts of ruminants, such as cows and sheep. Similarly, sulfate-reducing bacteria and Archaea, most of which are anaerobic ([link]), reduce sulfate to hydrogen sulfide to regenerate NAD^+ from NADH.



The green color seen in these coastal waters is from an eruption of hydrogen sulfide. Anaerobic, sulfate-reducing bacteria release hydrogen sulfide gas as they decompose algae in the water. (credit: NASA image courtesy Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC)

Note:
Concept in Action

Watch this video to see anaerobic cellular respiration in action.

Other fermentation methods occur in bacteria. Many prokaryotes are facultatively anaerobic. This means that they can switch between aerobic respiration and fermentation, depending on the availability of oxygen. Certain prokaryotes, like *Clostridia* bacteria, are obligate anaerobes. Obligate anaerobes live and grow in the absence of molecular oxygen. Oxygen is a poison to these microorganisms and kills them upon exposure. It should be noted that all forms of fermentation, except lactic acid fermentation, produce gas. The production of particular types of gas is used as an indicator of the fermentation of specific carbohydrates, which plays a role in the laboratory identification of the bacteria. The various methods of fermentation are used by different organisms to ensure an adequate supply of NAD^+ for the sixth step in glycolysis. Without these pathways, that step would not occur, and no ATP would be harvested from the breakdown of glucose.

Section Summary

If NADH cannot be metabolized through aerobic respiration, another electron acceptor is used. Most organisms will use some form of fermentation to accomplish the regeneration of NAD^+, ensuring the continuation of glycolysis. The regeneration of NAD^+ in fermentation is not accompanied by ATP production; therefore, the potential for NADH to produce ATP using an electron transport chain is not utilized.

Art Connections

Exercise:

Problem:

[link] Tremetol, a metabolic poison found in white snake root plant, prevents the metabolism of lactate. When cows eat this plant, Tremetol is concentrated in the milk. Humans who consume the milk become ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

Solution:

[link] The illness is caused by lactic acid build-up. Lactic acid levels rise after exercise, making the symptoms worse. Milk sickness is rare today, but was common in the Midwestern United States in the early 1800s.

Review Questions

Exercise:

Problem:

Which of the following fermentation methods can occur in animal skeletal muscles?

a. lactic acid fermentation
b. alcohol fermentation
c. mixed acid fermentation
d. propionic fermentation

Solution:

A

Free Response

Exercise:

Problem:

When muscle cells run out of oxygen, what happens to the potential for energy extraction from sugars and what pathways do the cell use?

Solution:

Without oxygen, oxidative phosphorylation and the citric acid cycle stop, so ATP is no longer generated through this mechanism, which extracts the greatest amount of energy from a sugar molecule. In addition, NADH accumulates, preventing glycolysis from going forward because of an absence of NAD^+. Lactic acid fermentation uses the electrons in NADH to generate lactic acid from pyruvate, which allows glycolysis to continue and thus a smaller amount of ATP can be generated by the cell.

Glossary

anaerobic cellular respiration
: the use of an electron acceptor other than oxygen to complete metabolism using electron transport-based chemiosmosis

fermentation
: the steps that follow the partial oxidation of glucose via glycolysis to regenerate NAD^+; occurs in the absence of oxygen and uses an organic compound as the final electron acceptor

Connections to Other Metabolic Pathways
By the end of this section, you will be able to:

- Discuss the way in which carbohydrate metabolic pathways, glycolysis, and the citric acid cycle interrelate with protein and lipid metabolic pathways
- Explain why metabolic pathways are not considered closed systems

You have learned about the catabolism of glucose, which provides energy to living cells. But living things consume more than just glucose for food. How does a turkey sandwich, which contains protein, provide energy to your cells? This happens because all of the catabolic pathways for carbohydrates, proteins, and lipids eventually connect into glycolysis and the citric acid cycle pathways ([link]). Metabolic pathways should be thought of as porous—that is, substances enter from other pathways, and other substances leave for other pathways. These pathways are not closed systems. Many of the products in a particular pathway are reactants in other pathways.

Connections of Other Sugars to Glucose Metabolism

Glycogen, a polymer of glucose, is a short-term energy storage molecule in animals. When there is adequate ATP present, excess glucose is converted into glycogen for storage. Glycogen is made and stored in the liver and muscle. Glycogen will be taken out of storage if blood sugar levels drop. The presence of glycogen in muscle cells as a source of glucose allows ATP to be produced for a longer time during exercise.

Sucrose is a disaccharide made from glucose and fructose bonded together. Sucrose is broken down in the small intestine, and the glucose and fructose are absorbed separately. Fructose is one of the three dietary monosaccharides, along with glucose and galactose (which is part of milk sugar, the disaccharide lactose), that are absorbed directly into the bloodstream during digestion. The catabolism of both fructose and galactose produces the same number of ATP molecules as glucose.

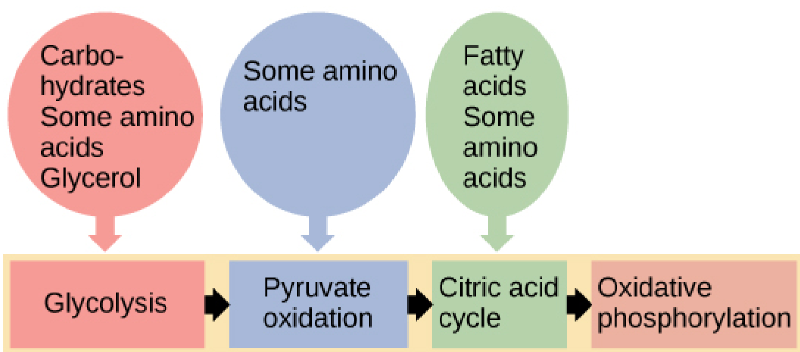
Connections of Proteins to Glucose Metabolism

Proteins are broken down by a variety of enzymes in cells. Most of the time, amino acids are recycled into new proteins. If there are excess amino acids, however, or if the body is in a state of famine, some amino acids will be shunted into pathways of glucose catabolism. Each amino acid must have its amino group removed prior to entry into these pathways. The amino group is converted into ammonia. In mammals, the liver synthesizes urea from two ammonia molecules and a carbon dioxide molecule. Thus, urea is the principal waste product in mammals from the nitrogen originating in amino acids, and it leaves the body in urine.

Connections of Lipids to Glucose Metabolism

The lipids that are connected to the glucose pathways are cholesterol and triglycerides. Cholesterol is a lipid that contributes to cell membrane flexibility and is a precursor of steroid hormones. The synthesis of cholesterol starts with acetyl CoA and proceeds in only one direction. The process cannot be reversed, and ATP is not produced.

Triglycerides are a form of long-term energy storage in animals. Triglycerides store about twice as much energy as carbohydrates. Triglycerides are made of glycerol and three fatty acids. Animals can make most of the fatty acids they need. Triglycerides can be both made and broken down through parts of the glucose catabolism pathways. Glycerol can be phosphorylated and proceeds through glycolysis. Fatty acids are broken into two-carbon units that enter the citric acid cycle.



Glycogen from the liver and muscles,

together with fats, can feed into the catabolic pathways for carbohydrates.

Note:

Evolution in Action

Pathways of Photosynthesis and Cellular Metabolism

Photosynthesis and cellular metabolism consist of several very complex pathways. It is generally thought that the first cells arose in an aqueous environment—a “soup” of nutrients. If these cells reproduced successfully and their numbers climbed steadily, it follows that the cells would begin to deplete the nutrients from the medium in which they lived, as they shifted the nutrients into their own cells. This hypothetical situation would have resulted in natural selection favoring those organisms that could exist by using the nutrients that remained in their environment and by manipulating these nutrients into materials that they could use to survive. Additionally, selection would favor those organisms that could extract maximal value from the available nutrients.

An early form of photosynthesis developed that harnessed the sun’s energy using compounds other than water as a source of hydrogen atoms, but this pathway did not produce free oxygen. It is thought that glycolysis developed prior to this time and could take advantage of simple sugars being produced, but these reactions were not able to fully extract the energy stored in the carbohydrates. A later form of photosynthesis used water as a source of hydrogen ions and generated free oxygen. Over time, the atmosphere became oxygenated. Living things adapted to exploit this new atmosphere and allowed respiration as we know it to evolve. When the full process of photosynthesis as we know it developed and the atmosphere became oxygenated, cells were finally able to use the oxygen expelled by photosynthesis to extract more energy from the sugar molecules using the citric acid cycle.

Section Summary

The breakdown and synthesis of carbohydrates, proteins, and lipids connect with the pathways of glucose catabolism. The carbohydrates that can also feed into glucose catabolism include galactose, fructose, and glycogen. These connect with glycolysis. The amino acids from proteins connect with glucose catabolism through pyruvate, acetyl CoA, and components of the citric acid cycle. Cholesterol synthesis starts with acetyl CoA, and the components of triglycerides are picked up by acetyl CoA and enter the citric acid cycle.

Multiple Choice

Exercise:

Problem:

The cholesterol synthesized by cells uses which component of the glycolytic pathway as a starting point?

a. glucose
b. acetyl CoA
c. pyruvate
d. carbon dioxide

Solution:

B

Exercise:

Problem: Beta oxidation is _________.

a. the breakdown of sugars
b. the assembly of sugars
c. the breakdown of fatty acids
d. the removal of amino groups from amino acids

Solution:

C

Free Response

Exercise:

Problem:

Would you describe metabolic pathways as inherently wasteful or inherently economical, and why?

Solution:

They are very economical. The substrates, intermediates, and products move between pathways and do so in response to finely tuned feedback inhibition loops that keep metabolism overall on an even keel. Intermediates in one pathway may occur in another, and they can move from one pathway to another fluidly in response to the needs of the cell.

Introduction
class="introduction"

This sage thrasher’s diet, like that of almost all organisms, depends on photosynthesis. (credit: modification of work by Dave Menke, U.S. Fish and Wildlife Service)

No matter how complex or advanced a machine, such as the latest cellular phone, the device cannot function without energy. Living things, similar to machines, have many complex components; they too cannot do anything without energy, which is why humans and all other organisms must “eat” in

some form or another. That may be common knowledge, but how many people realize that every bite of every meal ingested depends on the process of photosynthesis?

Overview of Photosynthesis
By the end of this section, you will be able to:

- Summarize the process of photosynthesis
- Explain the relevance of photosynthesis to other living things
- Identify the reactants and products of photosynthesis
- Describe the main structures involved in photosynthesis

All living organisms on earth consist of one or more cells. Each cell runs on the chemical energy found mainly in carbohydrate molecules (food), and the majority of these molecules are produced by one process: photosynthesis. Through photosynthesis, certain organisms convert solar energy (sunlight) into chemical energy, which is then used to build carbohydrate molecules. The energy used to hold these molecules together is released when an organism breaks down food. Cells then use this energy to perform work, such as cellular respiration.

The energy that is harnessed from photosynthesis enters the ecosystems of our planet continuously and is transferred from one organism to another. Therefore, directly or indirectly, the process of photosynthesis provides most of the energy required by living things on earth.

Photosynthesis also results in the release of oxygen into the atmosphere. In short, to eat and breathe, humans depend almost entirely on the organisms that carry out photosynthesis.

Note:
Concept in Action

Click the following link to learn more about photosynthesis.

Solar Dependence and Food Production

Some organisms can carry out photosynthesis, whereas others cannot. An **autotroph** is an organism that can produce its own food. The Greek roots of the word *autotroph* mean “self” (*auto*) “feeder” (*troph*). Plants are the best-known autotrophs, but others exist, including certain types of bacteria and algae ([link]). Oceanic algae contribute enormous quantities of food and oxygen to global food chains. Plants are also **photoautotrophs**, a type of autotroph that uses sunlight and carbon from carbon dioxide to synthesize chemical energy in the form of carbohydrates. All organisms carrying out photosynthesis require sunlight.

(a) Plants, (b) algae, and (c) certain bacteria, called cyanobacteria, are photoautotrophs that can carry out photosynthesis. Algae can grow over enormous areas in water, at times completely covering the surface. (credit a: Steve Hillebrand, U.S. Fish and Wildlife Service; credit b: "eutrophication&hypoxia"/Flickr; credit c: NASA; scale-bar data from Matt Russell)

Heterotrophs are organisms incapable of photosynthesis that must therefore obtain energy and carbon from food by consuming other organisms. The Greek roots of the word *heterotroph* mean “other” (*hetero*)

“feeder” (*troph*), meaning that their food comes from other organisms. Even if the food organism is another animal, this food traces its origins back to autotrophs and the process of photosynthesis. Humans are heterotrophs, as are all animals. Heterotrophs depend on autotrophs, either directly or indirectly. Deer and wolves are heterotrophs. A deer obtains energy by eating plants. A wolf eating a deer obtains energy that originally came from the plants eaten by that deer. The energy in the plant came from photosynthesis, and therefore it is the only autotroph in this example ([link]). Using this reasoning, all food eaten by humans also links back to autotrophs that carry out photosynthesis.

The energy stored in carbohydrate molecules from photosynthesis passes through the food chain. The predator that eats these deer is getting energy that originated in the photosynthetic vegetation that the deer consumed. (credit: Steve VanRiper, U.S. Fish and Wildlife Service)

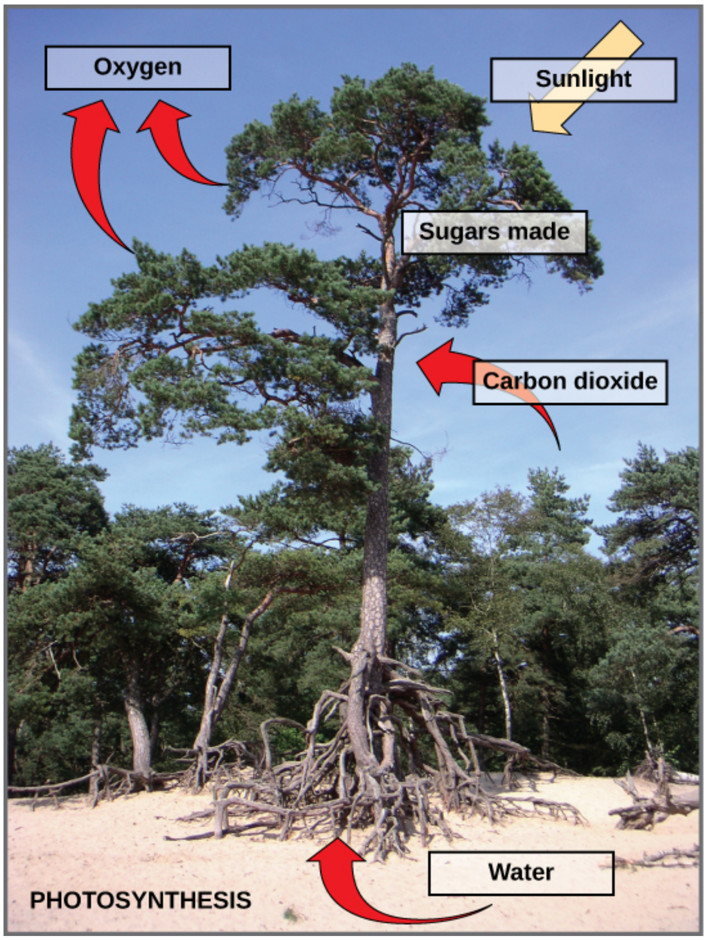
Note:
Biology in Action
Photosynthesis at the Grocery Store

Photosynthesis is the origin of the products that comprise the main elements of the human diet. (credit: Associação Brasileira de Supermercados)

Major grocery stores in the United States are organized into departments, such as dairy, meats, produce, bread, cereals, and so forth. Each aisle contains hundreds, if not thousands, of different products for customers to buy and consume ([link]).
Although there is a large variety, each item links back to photosynthesis. Meats and dairy products link to photosynthesis because the animals were fed plant-based foods. The breads, cereals, and pastas come largely from grains, which are the seeds of photosynthetic plants. What about desserts and drinks? All of these products contain sugar—the basic carbohydrate molecule produced directly from photosynthesis. The photosynthesis connection applies to every meal and every food a person consumes.

Main Structures and Summary of Photosynthesis

Photosynthesis requires sunlight, carbon dioxide, and water as starting reactants ([link]). After the process is complete, photosynthesis releases oxygen and produces carbohydrate molecules, most commonly glucose. These sugar molecules contain the energy that living things need to survive.



Photosynthesis uses solar energy, carbon dioxide, and water to release oxygen and to produce energy-storing sugar molecules.

The complex reactions of photosynthesis can be summarized by the chemical equation shown in [link].

Photosynthesis Equation			
Carbon dioxide +	Water $\xrightarrow{\text{SUNLIGHT}}$	Sugar +	Oxygen
$6CO_2$	$6H_2O$	$C_6H_{12}O_6$	$6O_2$

The process of photosynthesis can be represented by an equation, wherein carbon dioxide and water produce sugar and oxygen using energy from sunlight.

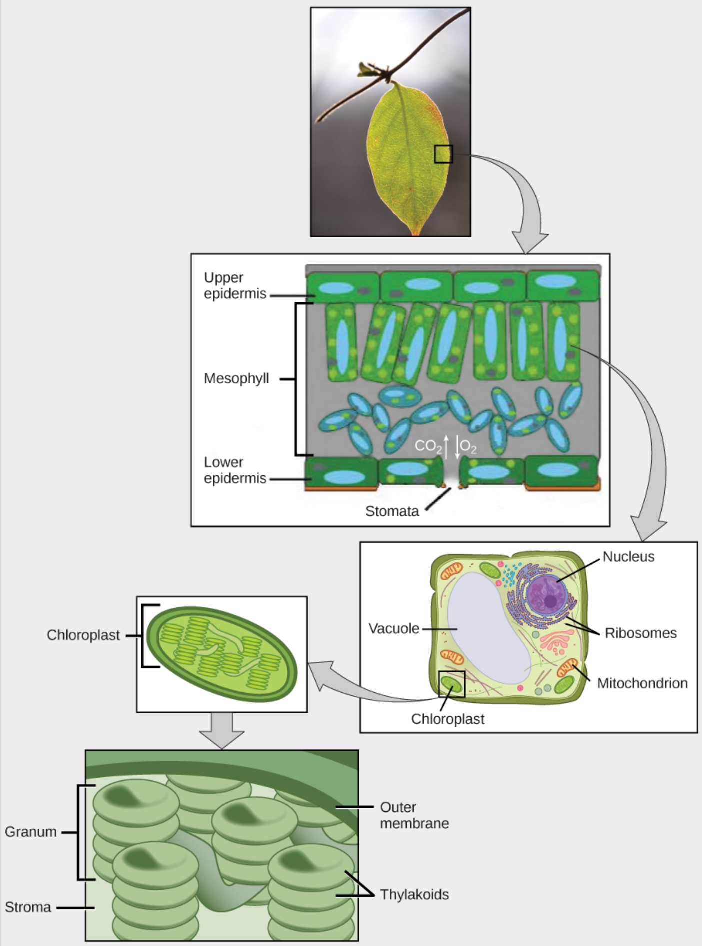
Although the equation looks simple, the many steps that take place during photosynthesis are actually quite complex, as in the way that the reaction summarizing cellular respiration represented many individual reactions. Before learning the details of how photoautotrophs turn sunlight into food, it is important to become familiar with the physical structures involved.

In plants, photosynthesis takes place primarily in leaves, which consist of many layers of cells and have differentiated top and bottom sides. The process of photosynthesis occurs not on the surface layers of the leaf, but rather in a middle layer called the **mesophyll** ([link]). The gas exchange of carbon dioxide and oxygen occurs through small, regulated openings called **stomata**.

In all autotrophic eukaryotes, photosynthesis takes place inside an organelle called a **chloroplast**. In plants, chloroplast-containing cells exist in the mesophyll. Chloroplasts have a double (inner and outer) membrane. Within the chloroplast is a third membrane that forms stacked, disc-shaped structures called **thylakoids**. Embedded in the thylakoid membrane are molecules of **chlorophyll**, a **pigment** (a molecule that absorbs light) through which the entire process of photosynthesis begins. Chlorophyll is responsible for the green color of plants. The thylakoid membrane encloses an internal space called the thylakoid space. Other types of pigments are also involved in photosynthesis, but chlorophyll is by far the most important. As shown in [link], a stack of thylakoids is called a **granum**, and

the space surrounding the granum is called **stroma** (not to be confused with stomata, the openings on the leaves).

Note:
Art Connection



Not all cells of a leaf carry out photosynthesis. Cells within the middle layer of a leaf have chloroplasts, which contain the photosynthetic apparatus. (credit "leaf": modification of work by Cory Zanker)

On a hot, dry day, plants close their stomata to conserve water. What impact will this have on photosynthesis?

The Two Parts of Photosynthesis

Photosynthesis takes place in two stages: the light-dependent reactions and the Calvin cycle. In the **light-dependent reactions**, which take place at the thylakoid membrane, chlorophyll absorbs energy from sunlight and then converts it into chemical energy with the use of water. The light-dependent reactions release oxygen from the hydrolysis of water as a byproduct. In the Calvin cycle, which takes place in the stroma, the chemical energy derived from the light-dependent reactions drives both the capture of carbon in carbon dioxide molecules and the subsequent assembly of sugar molecules. The two reactions use carrier molecules to transport the energy from one to the other. The carriers that move energy from the light-dependent reactions to the Calvin cycle reactions can be thought of as “full” because they bring energy. After the energy is released, the “empty” energy carriers return to the light-dependent reactions to obtain more energy.

Section Summary

The process of photosynthesis transformed life on earth. By harnessing energy from the sun, photosynthesis allowed living things to access enormous amounts of energy. Because of photosynthesis, living things gained access to sufficient energy, allowing them to evolve new structures and achieve the biodiversity that is evident today.

Only certain organisms, called autotrophs, can perform photosynthesis; they require the presence of chlorophyll, a specialized pigment that can absorb light and convert light energy into chemical energy. Photosynthesis uses carbon dioxide and water to assemble carbohydrate molecules (usually glucose) and releases oxygen into the air. Eukaryotic autotrophs, such as plants and algae, have organelles called chloroplasts in which photosynthesis takes place.

Art Connections

Exercise:

Problem:

[link] On a hot, dry day, plants close their stomata to conserve water. What impact will this have on photosynthesis?

Solution:

[link] Levels of carbon dioxide (a reactant) will fall, and levels of oxygen (a product) will rise. As a result, the rate of photosynthesis will slow down.

Multiple Choice

Exercise:

Problem: What two products result from photosynthesis?

a. water and carbon dioxide
b. water and oxygen
c. glucose and oxygen
d. glucose and carbon dioxide

Solution:

C

Exercise:

Problem:

Which statement about thylakoids in eukaryotes is *not* correct?

a. Thylakoids are assembled into stacks.
b. Thylakoids exist as a maze of folded membranes.

c. The space surrounding thylakoids is called stroma.
d. Thylakoids contain chlorophyll.

Solution:

B

Exercise:

Problem: From where does a heterotroph directly obtain its energy?

a. the sun
b. the sun and eating other organisms
c. eating other organisms
d. simple chemicals in the environment

Solution:

C

Free Response

Exercise:

Problem:

What is the overall purpose of the light reactions in photosynthesis?

Solution:

To convert solar energy into chemical energy that cells can use to do work.

Exercise:

Problem:

Why are carnivores, such as lions, dependent on photosynthesis to survive?

Solution:

Because lions eat animals that eat plants.

Glossary

autotroph
: an organism capable of producing its own food

chlorophyll
: the green pigment that captures the light energy that drives the reactions of photosynthesis

chloroplast
: the organelle where photosynthesis takes place

granum
: a stack of thylakoids located inside a chloroplast

heterotroph
: an organism that consumes other organisms for food

light-dependent reaction
: the first stage of photosynthesis where visible light is absorbed to form two energy-carrying molecules (ATP and NADPH)

mesophyll
: the middle layer of cells in a leaf

photoautotroph
: an organism capable of synthesizing its own food molecules (storing energy), using the energy of light

pigment
: a molecule that is capable of absorbing light energy

stoma
: the opening that regulates gas exchange and water regulation between leaves and the environment; plural: stomata

stroma
: the fluid-filled space surrounding the grana inside a chloroplast where the Calvin cycle reactions of photosynthesis take place

thylakoid
: a disc-shaped membranous structure inside a chloroplast where the light-dependent reactions of photosynthesis take place using chlorophyll embedded in the membranes

The Light-Dependent Reactions of Photosynthesis

By the end of this section, you will be able to:

- Explain how plants absorb energy from sunlight
- Describe how the wavelength of light affects its energy and color
- Describe how and where photosynthesis takes place within a plant

How can light be used to make food? It is easy to think of light as something that exists and allows living organisms, such as humans, to see, but light is a form of energy. Like all energy, light can travel, change form, and be harnessed to do work. In the case of photosynthesis, light energy is transformed into chemical energy, which autotrophs use to build carbohydrate molecules. However, autotrophs only use a specific component of sunlight ([link]).

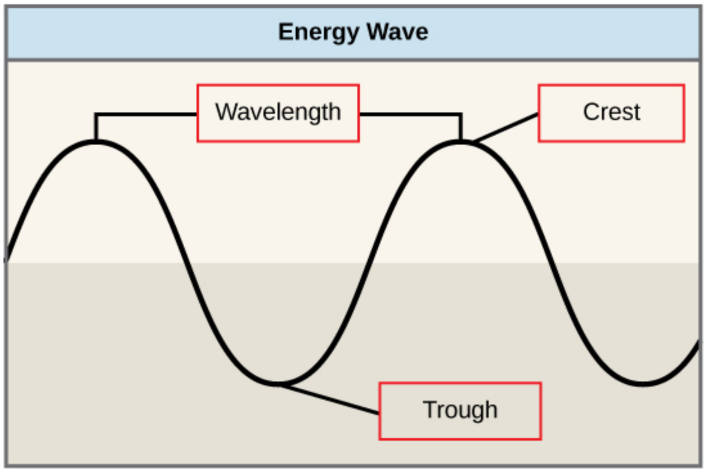
Autotrophs can capture light energy from the sun, converting it into chemical energy used to build food molecules. (credit: modification of work by Gerry Atwell, U.S. Fish and Wildlife Service)

Note:
Concept in Action

Watch the process of photosynthesis within a leaf in this video.

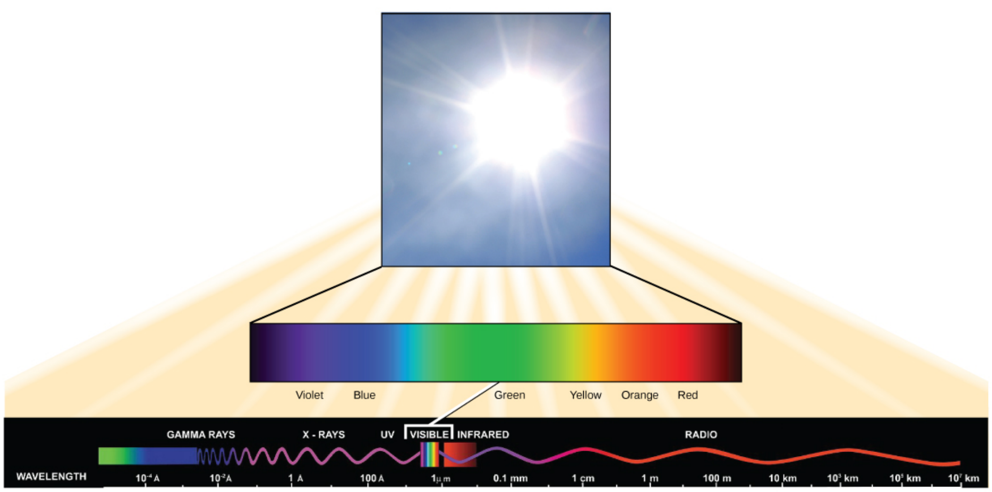
What Is Light Energy?

The sun emits an enormous amount of electromagnetic radiation (solar energy). Humans can see only a fraction of this energy, which is referred to as “visible light.” The manner in which solar energy travels can be described and measured as waves. Scientists can determine the amount of energy of a wave by measuring its **wavelength**, the distance between two consecutive, similar points in a series of waves, such as from crest to crest or trough to trough ([link]).



The wavelength of a single wave is the distance between two consecutive points along the wave.

Visible light constitutes only one of many types of electromagnetic radiation emitted from the sun. The **electromagnetic spectrum** is the range of all possible wavelengths of radiation ([link]). Each wavelength corresponds to a different amount of energy carried.



The sun emits energy in the form of electromagnetic radiation. This radiation exists in different wavelengths,

each of which has its own characteristic energy. Visible light is one type of energy emitted from the sun.

Each type of electromagnetic radiation has a characteristic range of wavelengths. The longer the wavelength (or the more stretched out it appears), the less energy is carried. Short, tight waves carry the most energy. This may seem illogical, but think of it in terms of a piece of moving rope. It takes little effort by a person to move a rope in long, wide waves. To make a rope move in short, tight waves, a person would need to apply significantly more energy.

The sun emits ([link]) a broad range of electromagnetic radiation, including X-rays and ultraviolet (UV) rays. The higher-energy waves are dangerous to living things; for example, X-rays and UV rays can be harmful to humans.

Absorption of Light

Light energy enters the process of photosynthesis when pigments absorb the light. In plants, pigment molecules absorb only visible light for photosynthesis. The visible light seen by humans as white light actually exists in a rainbow of colors. Certain objects, such as a prism or a drop of water, disperse white light to reveal these colors to the human eye. The visible light portion of the electromagnetic spectrum is perceived by the human eye as a rainbow of colors, with violet and blue having shorter wavelengths and, therefore, higher energy. At the other end of the spectrum toward red, the wavelengths are longer and have lower energy.

Understanding Pigments

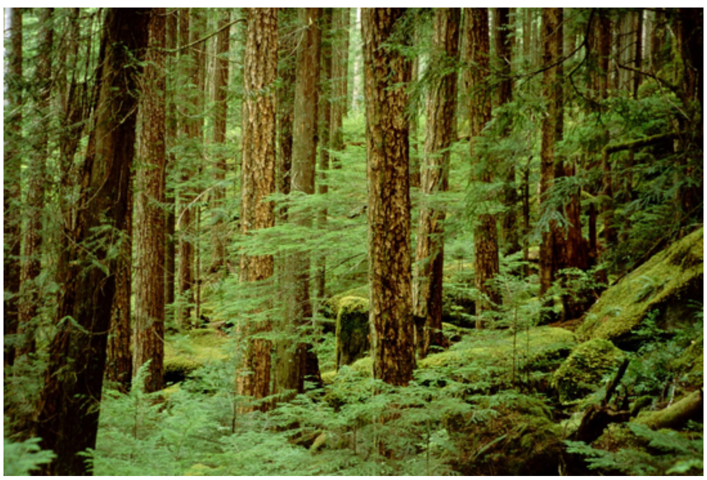
Different kinds of pigments exist, and each absorbs only certain wavelengths (colors) of visible light. Pigments reflect the color of the wavelengths that they cannot absorb.

All photosynthetic organisms contain a pigment called **chlorophyll *a***, which humans see as the common green color associated with plants.

Chlorophyll *a* absorbs wavelengths from either end of the visible spectrum (blue and red), but not from green. Because green is reflected, chlorophyll appears green.

Other pigment types include **chlorophyll *b*** (which absorbs blue and red-orange light) and the carotenoids. Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light, which is its **absorption spectrum**.

Many photosynthetic organisms have a mixture of pigments; between them, the organism can absorb energy from a wider range of visible-light wavelengths. Not all photosynthetic organisms have full access to sunlight. Some organisms grow underwater where light intensity decreases with depth, and certain wavelengths are absorbed by the water. Other organisms grow in competition for light. Plants on the rainforest floor must be able to absorb any bit of light that comes through, because the taller trees block most of the sunlight ([link]).



Plants that commonly grow in the shade benefit from having a variety of light-absorbing pigments. Each pigment can absorb different wavelengths of light, which allows the plant to absorb any light that passes

through the taller trees. (credit: Jason Hollinger)

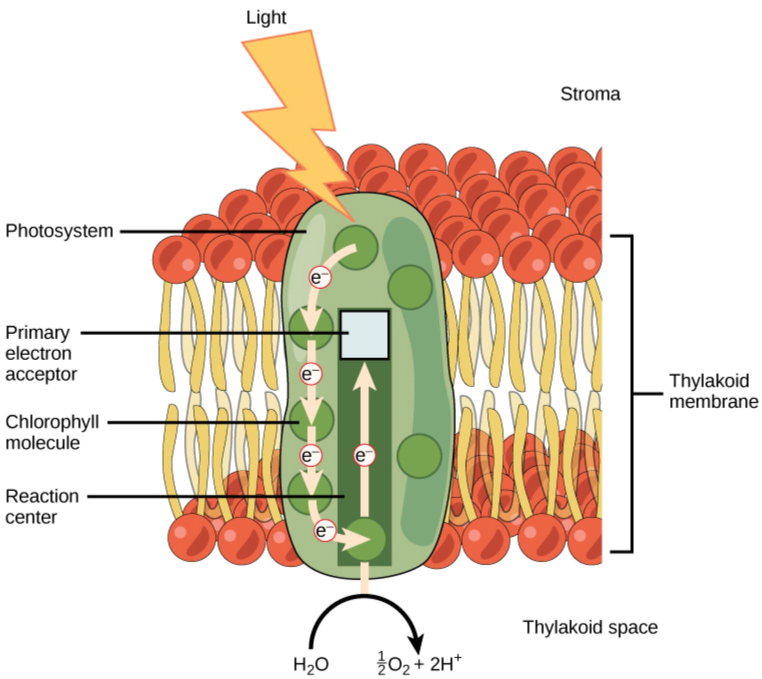
How Light-Dependent Reactions Work

The overall purpose of the light-dependent reactions is to convert light energy into chemical energy. This chemical energy will be used by the Calvin cycle to fuel the assembly of sugar molecules.

The light-dependent reactions begin in a grouping of pigment molecules and proteins called a **photosystem**. Photosystems exist in the membranes of thylakoids. A pigment molecule in the photosystem absorbs one **photon**, a quantity or “packet” of light energy, at a time.

A photon of light energy travels until it reaches a molecule of chlorophyll. The photon causes an electron in the chlorophyll to become “excited.” The energy given to the electron allows it to break free from an atom of the chlorophyll molecule. Chlorophyll is therefore said to “donate” an electron ([link]).

To replace the electron in the chlorophyll, a molecule of water is split. This splitting releases an electron and results in the formation of oxygen (O_2) and hydrogen ions (H^+) in the thylakoid space. Technically, each breaking of a water molecule releases a pair of electrons, and therefore can replace two donated electrons.

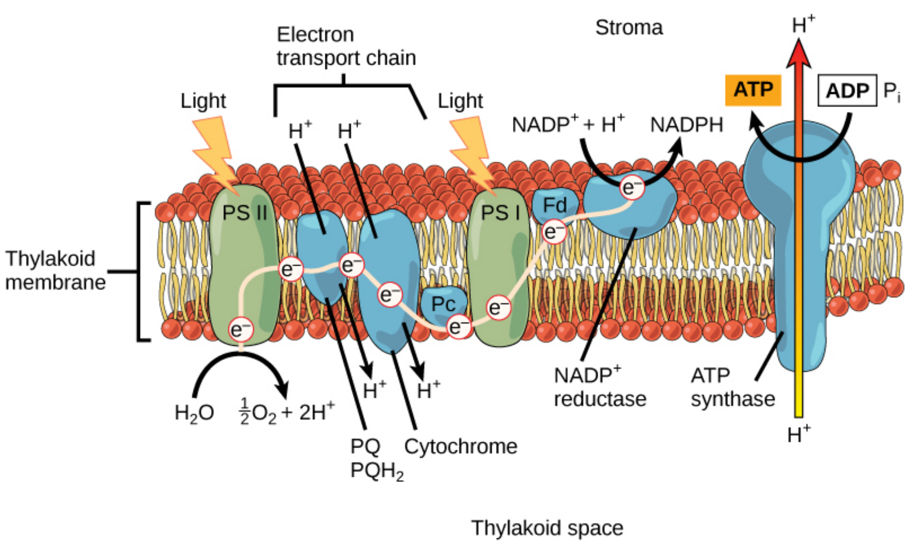


Light energy is absorbed by a chlorophyll molecule and is passed along a pathway to other chlorophyll molecules. The energy culminates in a molecule of chlorophyll found in the reaction center. The energy "excites" one of its electrons enough to leave the molecule and be transferred to a nearby primary electron acceptor. A molecule of water splits to release an electron, which is needed to replace the one donated. Oxygen and hydrogen ions are also formed from the splitting of water.

The replacing of the electron enables chlorophyll to respond to another photon. The oxygen molecules produced as byproducts find their way to the surrounding environment. The hydrogen ions play critical roles in the remainder of the light-dependent reactions.

Keep in mind that the purpose of the light-dependent reactions is to convert solar energy into chemical carriers that will be used in the Calvin cycle. In eukaryotes and some prokaryotes, two photosystems exist. The first is called photosystem II, which was named for the order of its discovery rather than for the order of the function.

After the photon hits, photosystem II transfers the free electron to the first in a series of proteins inside the thylakoid membrane called the electron transport chain. As the electron passes along these proteins, energy from the electron fuels membrane pumps that actively move hydrogen ions against their concentration gradient from the stroma into the thylakoid space. This is quite analogous to the process that occurs in the mitochondrion in which an electron transport chain pumps hydrogen ions from the mitochondrial stroma across the inner membrane and into the intermembrane space, creating an electrochemical gradient. After the energy is used, the electron is accepted by a pigment molecule in the next photosystem, which is called photosystem I ([link]).



From photosystem II, the electron travels along a series of proteins. This electron transport system uses the energy from the electron to pump hydrogen ions into the interior of the thylakoid. A

pigment molecule in photosystem I accepts the electron.

Generating an Energy Carrier: ATP

In the light-dependent reactions, energy absorbed by sunlight is stored by two types of energy-carrier molecules: ATP and NADPH. The energy that these molecules carry is stored in a bond that holds a single atom to the molecule. For ATP, it is a phosphate atom, and for NADPH, it is a hydrogen atom. Recall that NADH was a similar molecule that carried energy in the mitochondrion from the citric acid cycle to the electron transport chain. When these molecules release energy into the Calvin cycle, they each lose atoms to become the lower-energy molecules ADP and $NADP^+$.

The buildup of hydrogen ions in the thylakoid space forms an electrochemical gradient because of the difference in the concentration of protons (H^+) and the difference in the charge across the membrane that they create. This potential energy is harvested and stored as chemical energy in ATP through chemiosmosis, the movement of hydrogen ions down their electrochemical gradient through the transmembrane enzyme ATP synthase, just as in the mitochondrion.

The hydrogen ions are allowed to pass through the thylakoid membrane through an embedded protein complex called ATP synthase. This same protein generated ATP from ADP in the mitochondrion. The energy generated by the hydrogen ion stream allows ATP synthase to attach a third phosphate to ADP, which forms a molecule of ATP in a process called photophosphorylation. The flow of hydrogen ions through ATP synthase is called chemiosmosis, because the ions move from an area of high to low concentration through a semi-permeable structure.

Generating Another Energy Carrier: NADPH

The remaining function of the light-dependent reaction is to generate the other energy-carrier molecule, NADPH. As the electron from the electron transport chain arrives at photosystem I, it is re-energized with another

photon captured by chlorophyll. The energy from this electron drives the formation of NADPH from $NADP^+$ and a hydrogen ion (H^+). Now that the solar energy is stored in energy carriers, it can be used to make a sugar molecule.

Section Summary

In the first part of photosynthesis, the light-dependent reaction, pigment molecules absorb energy from sunlight. The most common and abundant pigment is chlorophyll *a*. A photon strikes photosystem II to initiate photosynthesis. Energy travels through the electron transport chain, which pumps hydrogen ions into the thylakoid space. This forms an electrochemical gradient. The ions flow through ATP synthase from the thylakoid space into the stroma in a process called chemiosmosis to form molecules of ATP, which are used for the formation of sugar molecules in the second stage of photosynthesis. Photosystem I absorbs a second photon, which results in the formation of an NADPH molecule, another energy carrier for the Calvin cycle reactions.

Multiple Choice

Exercise:

Problem:

What is the energy of a photon first used to do in photosynthesis?

a. split a water molecule
b. energize an electron
c. produce ATP
d. synthesize glucose

Solution:

B

Exercise:

Problem:

Which molecule absorbs the energy of a photon in photosynthesis?

a. ATP
b. glucose
c. chlorophyll
d. water

Solution:

C

Exercise:

Problem:

Plants produce oxygen when they photosynthesize. Where does the oxygen come from?

a. splitting water molecules
b. ATP synthesis
c. the electron transport chain
d. chlorophyll

Solution:

A

Exercise:

Problem: Which color(s) of light does chlorophyll *a* reflect?

a. red and blue
b. green
c. red
d. blue

Solution:

B

Free Response

Exercise:

Problem: Describe the pathway of energy in light-dependent reactions.

Solution:

The energy is present initially as light. A photon of light hits chlorophyll, causing an electron to be energized. The free electron travels through the electron transport chain, and the energy of the electron is used to pump hydrogen ions into the thylakoid space, transferring the energy into the electrochemical gradient. The energy of the electrochemical gradient is used to power ATP synthase, and the energy is transferred into a bond in the ATP molecule. In addition, energy from another photon can be used to create a high-energy bond in the molecule NADPH.

Glossary

absorption spectrum
: the specific pattern of absorption for a substance that absorbs electromagnetic radiation

chlorophyll *a*
: the form of chlorophyll that absorbs violet-blue and red light

chlorophyll *b*
: the form of chlorophyll that absorbs blue and red-orange light

electromagnetic spectrum

the range of all possible frequencies of radiation

photon
: a distinct quantity or "packet" of light energy

photosystem
: a group of proteins, chlorophyll, and other pigments that are used in the light-dependent reactions of photosynthesis to absorb light energy and convert it into chemical energy

wavelength
: the distance between consecutive points of a wave

The Calvin Cycle

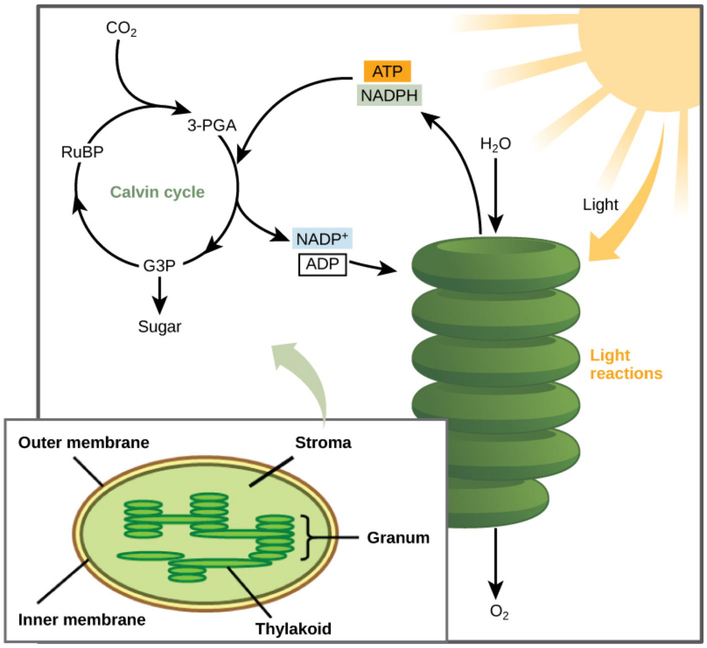
By the end of this section, you will be able to:

- Describe the Calvin cycle
- Define carbon fixation
- Explain how photosynthesis works in the energy cycle of all living organisms

After the energy from the sun is converted and packaged into ATP and NADPH, the cell has the fuel needed to build food in the form of carbohydrate molecules. The carbohydrate molecules made will have a backbone of carbon atoms. Where does the carbon come from? The carbon atoms used to build carbohydrate molecules comes from carbon dioxide, the gas that animals exhale with each breath. The **Calvin cycle** is the term used for the reactions of photosynthesis that use the energy stored by the light-dependent reactions to form glucose and other carbohydrate molecules.

The Interworkings of the Calvin Cycle

In plants, carbon dioxide (CO_2) enters the chloroplast through the stomata and diffuses into the stroma of the chloroplast—the site of the Calvin cycle reactions where sugar is synthesized. The reactions are named after the scientist who discovered them, and reference the fact that the reactions function as a cycle. Others call it the Calvin-Benson cycle to include the name of another scientist involved in its discovery ([link]).



Light-dependent reactions harness energy from the sun to produce ATP and NADPH. These energy-carrying molecules travel into the stroma where the Calvin cycle reactions take place.

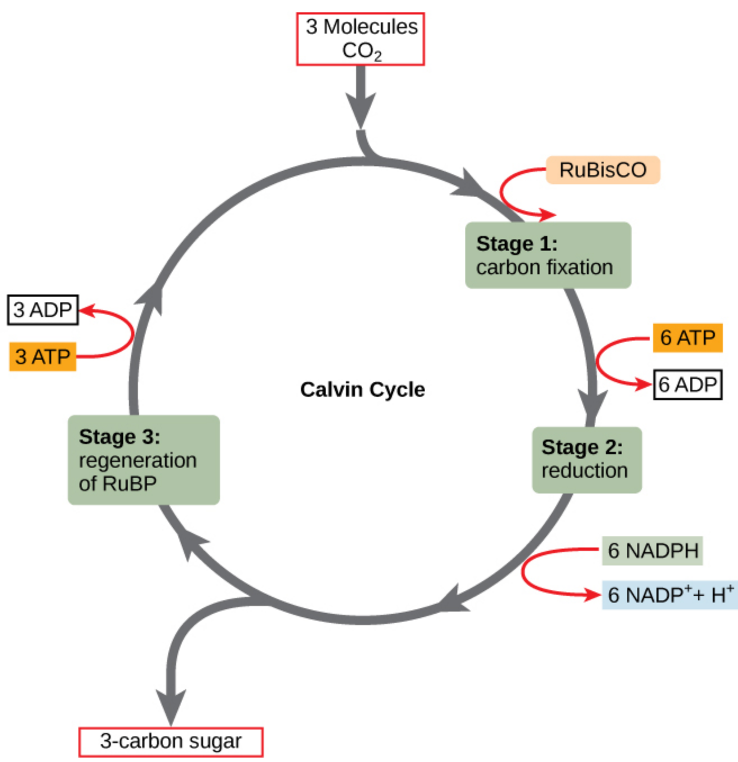
The Calvin cycle reactions ([link]) can be organized into three basic stages: fixation, reduction, and regeneration. In the stroma, in addition to CO_2, two other chemicals are present to initiate the Calvin cycle: an enzyme abbreviated RuBisCO, and the molecule ribulose bisphosphate (RuBP). RuBP has five atoms of carbon and a phosphate group on each end.

RuBisCO catalyzes a reaction between CO_2 and RuBP, which forms a six-carbon compound that is immediately converted into two three-carbon compounds. This process is called **carbon fixation**, because CO_2 is “fixed” from its inorganic form into organic molecules.

ATP and NADPH use their stored energy to convert the three-carbon compound, 3-PGA, into another three-carbon compound called G3P. This

type of reaction is called a reduction reaction, because it involves the gain of electrons. A reduction is the gain of an electron by an atom or molecule. The molecules of ADP and NAD^+, resulting from the reduction reaction, return to the light-dependent reactions to be re-energized.

One of the G3P molecules leaves the Calvin cycle to contribute to the formation of the carbohydrate molecule, which is commonly glucose ($C_6H_{12}O_6$). Because the carbohydrate molecule has six carbon atoms, it takes six turns of the Calvin cycle to make one carbohydrate molecule (one for each carbon dioxide molecule fixed). The remaining G3P molecules regenerate RuBP, which enables the system to prepare for the carbon-fixation step. ATP is also used in the regeneration of RuBP.



The Calvin cycle has three stages. In stage 1, the enzyme RuBisCO incorporates carbon dioxide into an organic molecule. In stage 2, the organic molecule is reduced. In stage 3, RuBP, the

molecule that starts the cycle, is regenerated so that the cycle can continue.

In summary, it takes six turns of the Calvin cycle to fix six carbon atoms from CO_2. These six turns require energy input from 12 ATP molecules and 12 NADPH molecules in the reduction step and 6 ATP molecules in the regeneration step.

Note:
Concept in Action

The following is a link to an animation of the Calvin cycle. Click Stage 1, Stage 2, and then Stage 3 to see G3P and ATP regenerate to form RuBP.

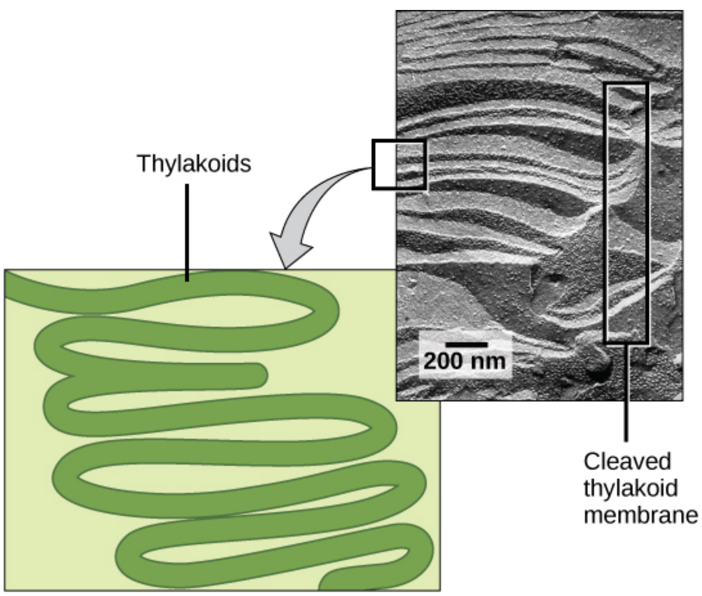
Note:
Evolution in Action
Photosynthesis
The shared evolutionary history of all photosynthetic organisms is conspicuous, as the basic process has changed little over eras of time. Even between the giant tropical leaves in the rainforest and tiny cyanobacteria, the process and components of photosynthesis that use water as an electron donor remain largely the same. Photosystems function to absorb light and use electron transport chains to convert energy. The Calvin cycle reactions assemble carbohydrate molecules with this energy.
However, as with all biochemical pathways, a variety of conditions leads to varied adaptations that affect the basic pattern. Photosynthesis in dry-

climate plants ([link]) has evolved with adaptations that conserve water. In the harsh dry heat, every drop of water and precious energy must be used to survive. Two adaptations have evolved in such plants. In one form, a more efficient use of CO_2 allows plants to photosynthesize even when CO_2 is in short supply, as when the stomata are closed on hot days. The other adaptation performs preliminary reactions of the Calvin cycle at night, because opening the stomata at this time conserves water due to cooler temperatures. In addition, this adaptation has allowed plants to carry out low levels of photosynthesis without opening stomata at all, an extreme mechanism to face extremely dry periods.

Living in the harsh conditions of the desert has led plants like this cactus to evolve variations in reactions outside the Calvin cycle. These variations increase efficiency and help conserve water and energy. (credit: Piotr Wojtkowski)

Photosynthesis in Prokaryotes

The two parts of photosynthesis—the light-dependent reactions and the Calvin cycle—have been described, as they take place in chloroplasts. However, prokaryotes, such as cyanobacteria, lack membrane-bound organelles. Prokaryotic photosynthetic autotrophic organisms have infoldings of the plasma membrane for chlorophyll attachment and photosynthesis ([link]). It is here that organisms like cyanobacteria can carry out photosynthesis.



A photosynthetic prokaryote has infolded regions of the plasma membrane that function like thylakoids. Although these are not contained in an organelle, such as a chloroplast, all of the necessary components are present to carry out

photosynthesis. (credit: scale-bar data from Matt Russell)

The Energy Cycle

Living things access energy by breaking down carbohydrate molecules. However, if plants make carbohydrate molecules, why would they need to break them down? Carbohydrates are storage molecules for energy in all living things. Although energy can be stored in molecules like ATP, carbohydrates are much more stable and efficient reservoirs for chemical energy. Photosynthetic organisms also carry out the reactions of respiration to harvest the energy that they have stored in carbohydrates, for example, plants have mitochondria in addition to chloroplasts.

You may have noticed that the overall reaction for photosynthesis:
Equation:

$$6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$$

is the reverse of the overall reaction for cellular respiration:
Equation:

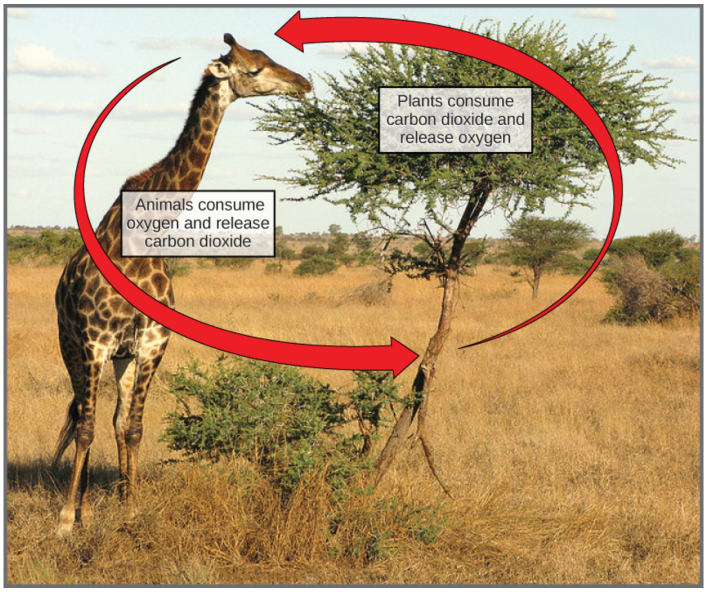
$$6O_2 + C_6H_{12}O_6 \rightarrow 6CO_2 + 6H_2O$$

Photosynthesis produces oxygen as a byproduct, and respiration produces carbon dioxide as a byproduct.

In nature, there is no such thing as waste. Every single atom of matter is conserved, recycling indefinitely. Substances change form or move from one type of molecule to another, but never disappear ([link]).

CO_2 is no more a form of waste produced by respiration than oxygen is a waste product of photosynthesis. Both are byproducts of reactions that move on to other reactions. Photosynthesis absorbs energy to build carbohydrates in chloroplasts, and aerobic cellular respiration releases

energy by using oxygen to break down carbohydrates. Both organelles use electron transport chains to generate the energy necessary to drive other reactions. Photosynthesis and cellular respiration function in a biological cycle, allowing organisms to access life-sustaining energy that originates millions of miles away in a star.



In the carbon cycle, the reactions of photosynthesis and cellular respiration share reciprocal reactants and products. (credit: modification of work by Stuart Bassil)

Section Summary

Using the energy carriers formed in the first stage of photosynthesis, the Calvin cycle reactions fix CO_2 from the environment to build carbohydrate molecules. An enzyme, RuBisCO, catalyzes the fixation reaction, by combining CO_2 with RuBP. The resulting six-carbon compound is broken down into two three-carbon compounds, and the energy in ATP and

NADPH is used to convert these molecules into G3P. One of the three-carbon molecules of G3P leaves the cycle to become a part of a carbohydrate molecule. The remaining G3P molecules stay in the cycle to be formed back into RuBP, which is ready to react with more CO_2. Photosynthesis forms a balanced energy cycle with the process of cellular respiration. Plants are capable of both photosynthesis and cellular respiration, since they contain both chloroplasts and mitochondria.

Multiple Choice

Exercise:

Problem: Where in plant cells does the Calvin cycle take place?

a. thylakoid membrane
b. thylakoid space
c. stroma
d. granum

Solution:

C

Exercise:

Problem: Which statement correctly describes carbon fixation?

a. the conversion of CO_2 to an organic compound
b. the use of RuBisCO to form 3-PGA
c. the production of carbohydrate molecules from G3P
d. the formation of RuBP from G3P molecules
e. the use of ATP and NADPH to reduce CO_2

Solution:

A

Exercise:

Problem:

What is the molecule that leaves the Calvin cycle to be converted into glucose?

a. ADP
b. G3P
c. RuBP
d. 3-PGA

Solution:

B

Free Response

Exercise:

Problem:

Which part of the Calvin cycle would be affected if a cell could not produce the enzyme RuBisCO?

Solution:

None of the cycle could take place, because RuBisCO is essential in fixing carbon dioxide. Specifically, RuBisCO catalyzes the reaction between carbon dioxide and RuBP at the start of the cycle.

Exercise:

Problem:

Explain the reciprocal nature of the net chemical reactions for photosynthesis and respiration.

Solution:

Photosynthesis takes the energy of sunlight and combines water and carbon dioxide to produce sugar and oxygen as a waste product. The reactions of respiration take sugar and consume oxygen to break it down into carbon dioxide and water, releasing energy. Thus, the reactants of photosynthesis are the products of respiration, and vice versa.

Glossary

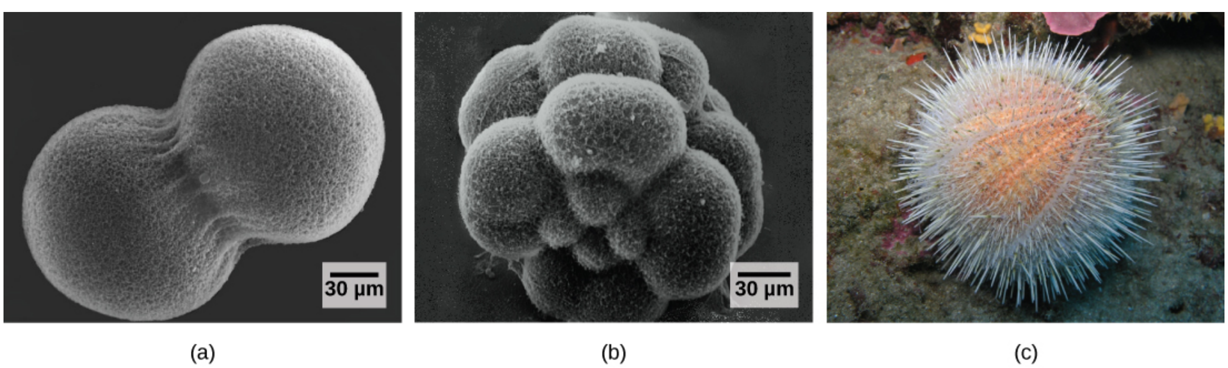
Calvin cycle
: the reactions of photosynthesis that use the energy stored by the light-dependent reactions to form glucose and other carbohydrate molecules

carbon fixation
: the process of converting inorganic CO_2 gas into organic compounds

Introduction
class="introduction"

A sea urchin begins life as a single cell that (a) divides to form two cells, visible by scanning electron microscopy. After four rounds of cell division, (b) there are 16 cells, as seen in this SEM image. After many rounds of cell division, the individual develops into a complex, multicellular organism, as seen in this (c) mature sea urchin. (credit a: modificatio

n of work by Evelyn Spiegel, Louisa Howard; credit b: modification of work by Evelyn Spiegel, Louisa Howard; credit c: modification of work by Marco Busdraghi; scale-bar data from Matt Russell)



The individual sexually reproducing organism—including humans—begins life as a fertilized egg, or zygote. Trillions of cell divisions subsequently occur in a controlled manner to produce a complex, multicellular human. In other words, that original single cell was the ancestor of every other cell in the body. Once a human individual is fully grown, cell reproduction is still

necessary to repair or regenerate tissues. For example, new blood and skin cells are constantly being produced. All multicellular organisms use cell division for growth, and in most cases, the maintenance and repair of cells and tissues. Single-celled organisms use cell division as their method of reproduction.

The Genome

By the end of this section, you will be able to:

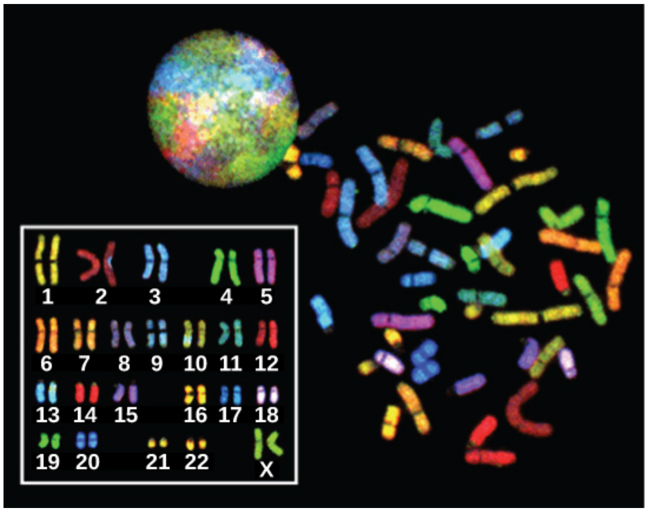
- Describe the prokaryotic and eukaryotic genome
- Distinguish between chromosomes, genes, and traits

The continuity of life from one cell to another has its foundation in the reproduction of cells by way of the cell cycle. The cell cycle is an orderly sequence of events in the life of a cell from the division of a single parent cell to produce two new daughter cells, to the subsequent division of those daughter cells. The mechanisms involved in the cell cycle are highly conserved across eukaryotes. Organisms as diverse as protists, plants, and animals employ similar steps.

Genomic DNA

Before discussing the steps a cell undertakes to replicate, a deeper understanding of the structure and function of a cell's genetic information is necessary. A cell's complete complement of DNA is called its **genome**. In prokaryotes, the genome is composed of a single, double-stranded DNA molecule in the form of a loop or circle. The region in the cell containing this genetic material is called a nucleoid. Some prokaryotes also have smaller loops of DNA called plasmids that are not essential for normal growth.

In eukaryotes, the genome comprises several double-stranded, linear DNA molecules ([link]) bound with proteins to form complexes called chromosomes. Each species of eukaryote has a characteristic number of chromosomes in the nuclei of its cells. Human body cells (somatic cells) have 46 chromosomes. A somatic cell contains two matched sets of chromosomes, a configuration known as **diploid**. The letter *n* is used to represent a single set of chromosomes; therefore a diploid organism is designated 2*n*. Human cells that contain one set of 23 chromosomes are called **gametes**, or sex cells; these eggs and sperm are designated *n*, or **haploid**.



There are 23 pairs of homologous chromosomes in a female human somatic cell. These chromosomes are viewed within the nucleus (top), removed from a cell in mitosis (right), and arranged according to length (left) in an arrangement called a karyotype. In this image, the chromosomes were exposed to fluorescent stains to distinguish them. (credit: “718 Bot”/Wikimedia Commons, National Human Genome Research)

The matched pairs of chromosomes in a diploid organism are called **homologous chromosomes**. Homologous chromosomes are the same length and have specific nucleotide segments called **genes** in exactly the same location, or **locus**. Genes, the functional units of chromosomes, determine specific characteristics by coding for specific proteins. Traits are the different forms of a characteristic. For example, the shape of earlobes is a characteristic with traits of free or attached.

Each copy of the homologous pair of chromosomes originates from a different parent; therefore, the copies of each of the genes themselves may not be identical. The variation of individuals within a species is caused by the specific combination of the genes inherited from both parents. For example, there are three possible gene sequences on the human chromosome that codes for blood type: sequence A, sequence B, and sequence O. Because all diploid human cells have two copies of the chromosome that determines blood type, the blood type (the trait) is determined by which two versions of the marker gene are inherited. It is possible to have two copies of the same gene sequence, one on each homologous chromosome (for example, AA, BB, or OO), or two different sequences, such as AB.

Minor variations in traits such as those for blood type, eye color, and height contribute to the natural variation found within a species. The sex chromosomes, X and Y, are the single exception to the rule of homologous chromosomes; other than a small amount of homology that is necessary to reliably produce gametes, the genes found on the X and Y chromosomes are not the same.

Section Summary

Prokaryotes have a single loop chromosome, whereas eukaryotes have multiple, linear chromosomes surrounded by a nuclear membrane. Human somatic cells have 46 chromosomes consisting of two sets of 22 homologous chromosomes and a pair of nonhomologous sex chromosomes. This is the 2*n*, or diploid, state. Human gametes have 23 chromosomes or one complete set of chromosomes. This is the *n*, or haploid, state. Genes are segments of DNA that code for a specific protein or RNA molecule. An organism's traits are determined in large part by the genes inherited from each parent, but also by the environment that they experience. Genes are expressed as characteristics of the organism and each characteristic may have different variants called traits that are caused by differences in the DNA sequence for a gene.

Multiple Choice

Exercise:

Problem:

A diploid cell has _________ the number of chromosomes as a haploid cell.

a. one-fourth
b. one-half
c. twice
d. four times

Solution:

C

Exercise:

Problem:

An organism’s traits are determined by the specific combination of inherited _________.

a. cells
b. genes
c. proteins
d. chromatids

Solution:

B

Free Response

Exercise:

Problem:

Compare and contrast a human somatic cell to a human gamete.

Solution:

Human somatic cells have 46 chromosomes, including 22 homologous pairs and one pair of nonhomologous sex chromosomes. This is the $2n$, or diploid, condition. Human gametes have 23 chromosomes, one each of 23 unique chromosomes. This is the n, or haploid, condition.

Glossary

diploid
: describes a cell, nucleus, or organism containing two sets of chromosomes ($2n$)

gamete
: a haploid reproductive cell or sex cell (sperm or egg)

gene
: the physical and functional unit of heredity; a sequence of DNA that codes for a specific peptide or RNA molecule

genome
: the entire genetic complement (DNA) of an organism

haploid
: describes a cell, nucleus, or organism containing one set of chromosomes (n)

homologous chromosomes
: chromosomes of the same length with genes in the same location; diploid organisms have pairs of homologous chromosomes, and the members of each pair come from different parents

locus

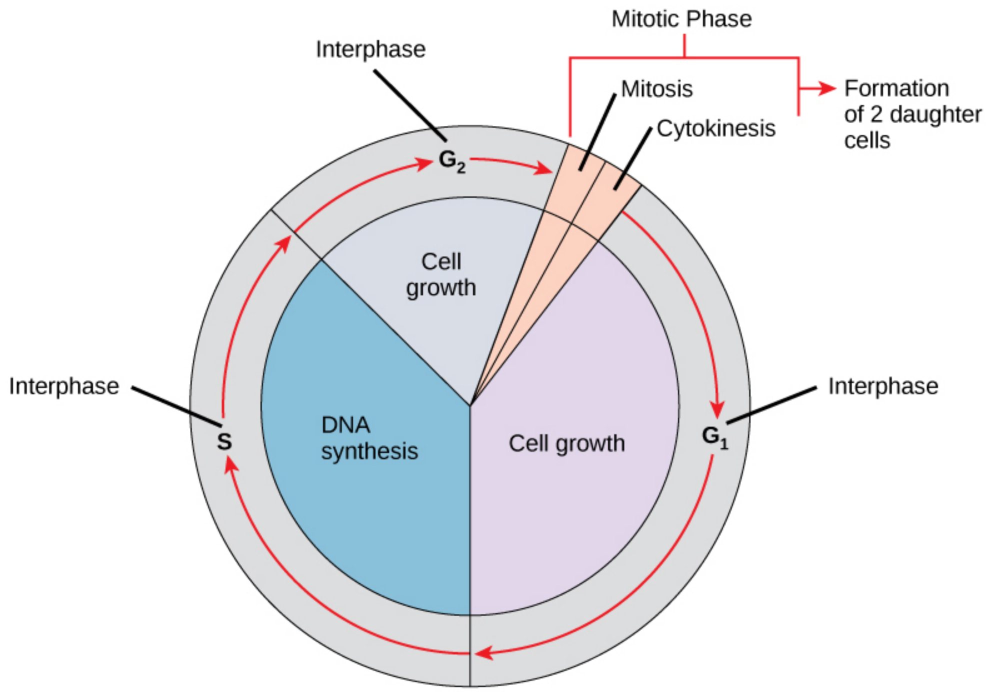
the position of a gene on a chromosome

The Cell Cycle

By the end of this section, you will be able to:

- Describe the three stages of interphase
- Discuss the behavior of chromosomes during mitosis and how the cytoplasmic content divides during cytokinesis
- Define the quiescent G_0 phase
- Explain how the three internal control checkpoints occur at the end of G_1, at the G_2–M transition, and during metaphase

The **cell cycle** is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and division that produce two genetically identical cells. The cell cycle has two major phases: interphase and the mitotic phase ([link]). During **interphase**, the cell grows and DNA is replicated. During the **mitotic phase**, the replicated DNA and cytoplasmic contents are separated and the cell divides. Watch this video about the cell cycle: https://www.youtube.com/watch?v=Wy3N5NCZBHQ



A cell moves through a series of phases in an orderly manner. During interphase, G_1 involves cell growth and

protein synthesis, the S phase involves DNA replication and the replication of the centrosome, and G_2 involves further growth and protein synthesis. The mitotic phase follows interphase. Mitosis is nuclear division during which duplicated chromosomes are segregated and distributed into daughter nuclei. Usually the cell will divide after mitosis in a process called cytokinesis in which the cytoplasm is divided and two daughter cells are formed.

Interphase

During interphase, the cell undergoes normal processes while also preparing for cell division. For a cell to move from interphase to the mitotic phase, many internal and external conditions must be met. The three stages of interphase are called G_1, S, and G_2.

G_1 Phase

The first stage of interphase is called the **G_1 phase**, or first gap, because little change is visible. However, during the G_1 stage, the cell is quite active at the biochemical level. The cell is accumulating the building blocks of chromosomal DNA and the associated proteins, as well as accumulating enough energy reserves to complete the task of replicating each chromosome in the nucleus.

S Phase

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In the **S phase** (synthesis phase), DNA replication results in the formation of two identical copies of each chromosome—sister chromatids—that are firmly attached at the centromere region. At this stage,

each chromosome is made of two sister chromatids and is a duplicated chromosome. The centrosome is duplicated during the S phase. The two centrosomes will give rise to the **mitotic spindle**, the apparatus that orchestrates the movement of chromosomes during mitosis. The centrosome consists of a pair of rod-like **centrioles** at right angles to each other. Centrioles help organize cell division. Centrioles are not present in the centrosomes of many eukaryotic species, such as plants and most fungi.

G_2 Phase

In the **G_2 phase**, or second gap, the cell replenishes its energy stores and synthesizes the proteins necessary for chromosome manipulation. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic spindle. There may be additional cell growth during G_2. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.

The Mitotic Phase

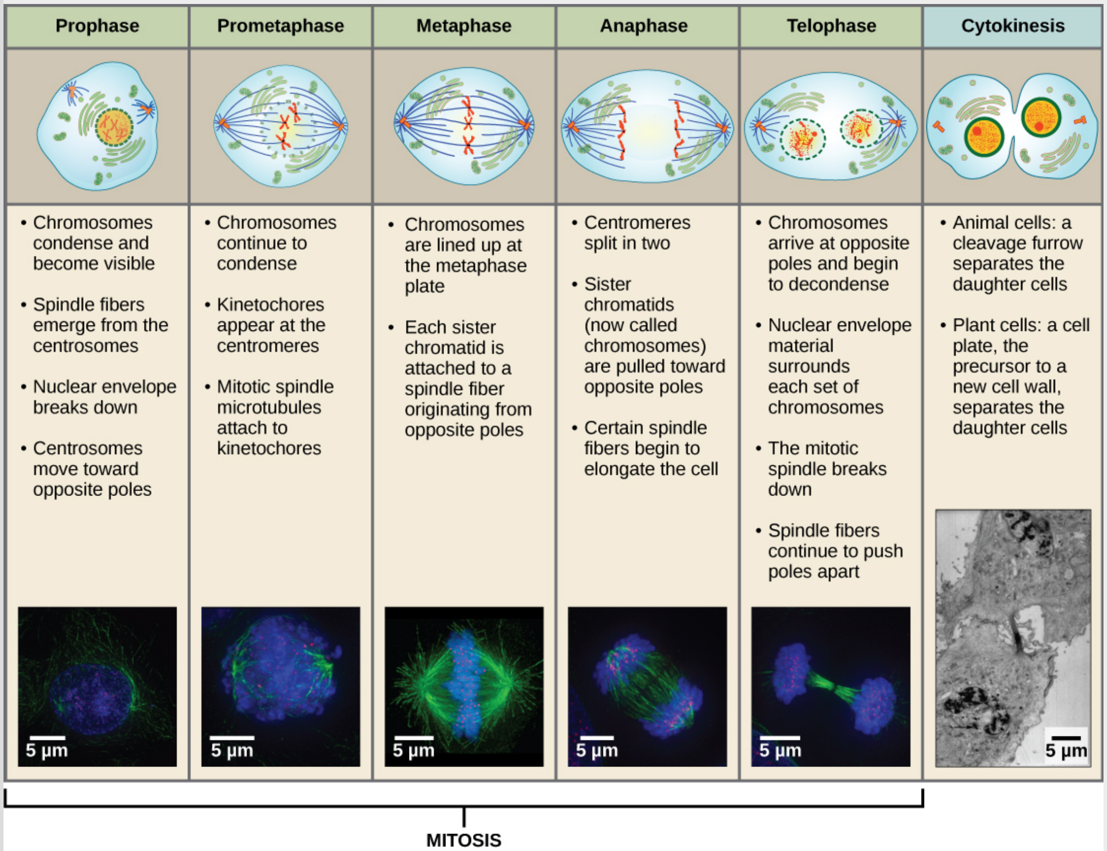
To make two daughter cells, the contents of the nucleus and the cytoplasm must be divided. The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and moved to opposite poles of the cell, and then the cell is divided into two new identical daughter cells. The first portion of the mitotic phase, **mitosis**, is composed of five stages, which accomplish nuclear division. The second portion of the mitotic phase, called cytokinesis, is the physical separation of the cytoplasmic components into two daughter cells.

Mitosis

Mitosis is divided into a series of phases—prophase, prometaphase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus ([link]).

Note:

Art Connection



Animal cell mitosis is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase—visualized here by light microscopy with fluorescence. Mitosis is usually accompanied by cytokinesis, shown here by a transmission electron microscope. (credit "diagrams": modification of work by Mariana Ruiz Villareal; credit "mitosis micrographs": modification of work by Roy van Heesbeen; credit "cytokinesis micrograph": modification of work by the Wadsworth Center, NY State Department of Health; donated to the Wikimedia foundation; scale-bar data from Matt Russell)

Which of the following is the correct order of events in mitosis?

a. Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus re-forms and the cell divides. The sister chromatids separate.
b. The kinetochore becomes attached to the mitotic spindle. The sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus re-forms and the cell divides.
c. The kinetochore becomes attached to metaphase plate. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus re-forms and the cell divides.
d. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. The kinetochore breaks apart and the sister chromatids separate. The nucleus re-forms and the cell divides.

During **prophase**, the "first phase," several events must occur to provide access to the chromosomes in the nucleus. The nuclear envelope starts to break into small vesicles, and the Golgi apparatus and endoplasmic reticulum fragment and disperse to the periphery of the cell. The nucleolus disappears. The centrosomes begin to move to opposite poles of the cell. The microtubules that form the basis of the mitotic spindle extend between the centrosomes, pushing them farther apart as the microtubule fibers lengthen. The sister chromatids begin to coil more tightly and become visible under a light microscope.

During **prometaphase**, many processes that were begun in prophase continue to advance and culminate in the formation of a connection between the chromosomes and cytoskeleton. The remnants of the nuclear envelope disappear. The mitotic spindle continues to develop as more microtubules assemble and stretch across the length of the former nuclear area. Chromosomes become more condensed and visually discrete. Each sister chromatid attaches to spindle microtubules at the centromere via a protein complex called the **kinetochore**.

During **metaphase**, all of the chromosomes are aligned in a plane called the **metaphase plate**, or the equatorial plane, midway between the two poles of the cell. The sister chromatids are still tightly attached to each other. At this time, the chromosomes are maximally condensed.

During **anaphase**, the sister chromatids at the equatorial plane are split apart at the centromere. Each chromatid, now called a chromosome, is pulled rapidly toward the centrosome to which its microtubule was attached. The cell becomes visibly elongated as the non-kinetochore microtubules slide against each other at the metaphase plate where they overlap.

During **telophase**, all of the events that set up the duplicated chromosomes for mitosis during the first three phases are reversed. The chromosomes reach the opposite poles and begin to decondense (unravel). The mitotic spindles are broken down into monomers that will be used to assemble cytoskeleton components for each daughter cell. Nuclear envelopes form around chromosomes.

Note:
Concept in Action

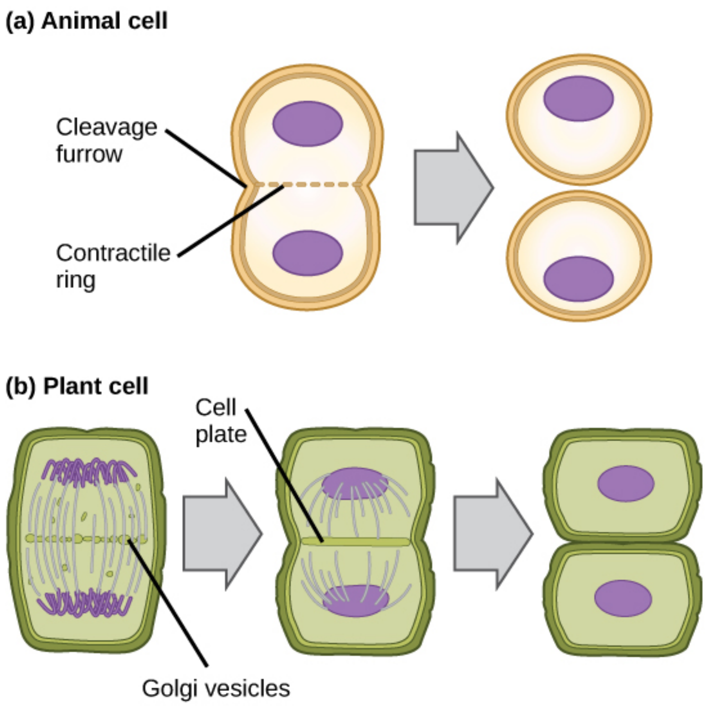
This page of movies illustrates different aspects of mitosis. Watch the movie entitled "DIC microscopy of cell division in a newt lung cell" and identify the phases of mitosis.

Cytokinesis

Cytokinesis is the second part of the mitotic phase during which cell division is completed by the physical separation of the cytoplasmic components into two daughter cells. Although the stages of mitosis are similar for most eukaryotes, the process of cytokinesis is quite different for eukaryotes that have cell walls, such as plant cells.

In cells such as animal cells that lack cell walls, cytokinesis begins following the onset of anaphase. A contractile ring composed of actin filaments forms just inside the plasma membrane at the former metaphase plate. The actin filaments pull the equator of the cell inward, forming a fissure. This fissure, or "crack," is called the **cleavage furrow**. The furrow deepens as the actin ring contracts, and eventually the membrane and cell are cleaved in two ([link]).

In plant cells, a cleavage furrow is not possible because of the rigid cell walls surrounding the plasma membrane. A new cell wall must form between the daughter cells. During interphase, the Golgi apparatus accumulates enzymes, structural proteins, and glucose molecules prior to breaking up into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles move on microtubules to collect at the metaphase plate. There, the vesicles fuse from the center toward the cell walls; this structure is called a **cell plate**. As more vesicles fuse, the cell plate enlarges until it merges with the cell wall at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall of cellulose. The Golgi membranes become the plasma membrane on either side of the new cell wall ([link]).

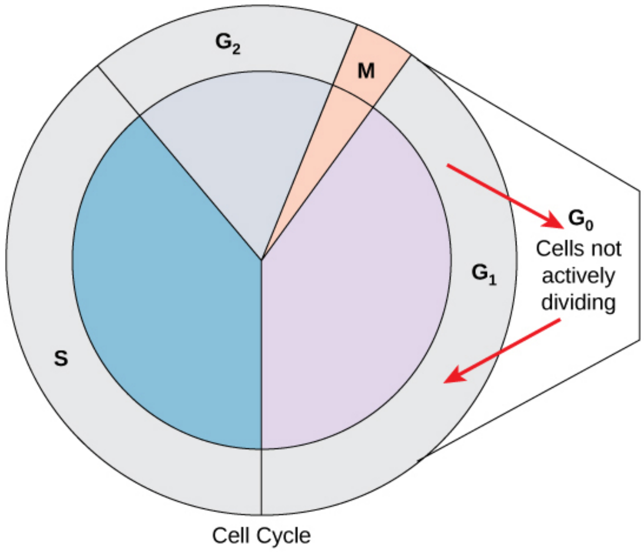


In part (a), a cleavage furrow forms at the former metaphase plate in the animal cell. The plasma membrane is drawn in by a ring of actin fibers contracting just inside the membrane. The cleavage furrow deepens until the cells are pinched in two. In part (b), Golgi vesicles coalesce at the former metaphase plate in a plant cell. The vesicles fuse and form the cell plate. The cell plate grows from the center toward the cell walls. New cell walls are made from the vesicle contents.

G_0 Phase

Not all cells adhere to the classic cell-cycle pattern in which a newly formed daughter cell immediately enters interphase, closely followed by the

mitotic phase. Cells in the **G_0 phase** are not actively preparing to divide. The cell is in a quiescent (inactive) stage, having exited the cell cycle. Some cells enter G_0 temporarily until an external signal triggers the onset of G_1. Other cells that never or rarely divide, such as mature cardiac muscle and nerve cells, remain in G_0 permanently ([link]).



Cells that are not actively preparing to divide enter an alternate phase called G_0. In some cases, this is a temporary condition until triggered to enter G_1. In other cases, the cell will remain in G_0 permanently.

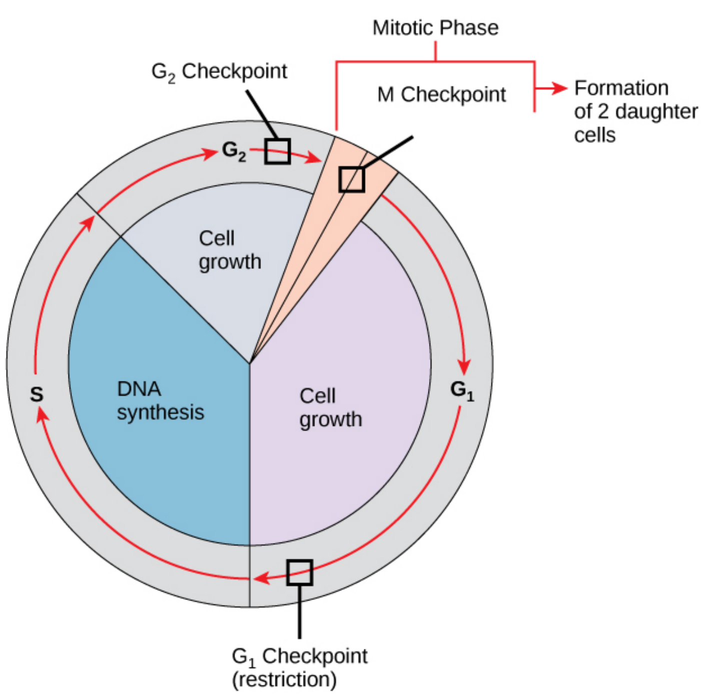
Control of the Cell Cycle

The length of the cell cycle is highly variable even within the cells of an individual organism. In humans, the frequency of cell turnover ranges from a few hours in early embryonic development to an average of two to five

days for epithelial cells, or to an entire human lifetime spent in G_0 by specialized cells such as cortical neurons or cardiac muscle cells. There is also variation in the time that a cell spends in each phase of the cell cycle. When fast-dividing mammalian cells are grown in culture (outside the body under optimal growing conditions), the length of the cycle is approximately 24 hours. In rapidly dividing human cells with a 24-hour cell cycle, the G_1 phase lasts approximately 11 hours. The timing of events in the cell cycle is controlled by mechanisms that are both internal and external to the cell.

Regulation at Internal Checkpoints

It is essential that daughter cells be exact duplicates of the parent cell. Mistakes in the duplication or distribution of the chromosomes lead to mutations that may be passed forward to every new cell produced from the abnormal cell. To prevent a compromised cell from continuing to divide, there are internal control mechanisms that operate at three main **cell cycle checkpoints** at which the cell cycle can be stopped until conditions are favorable. These checkpoints occur near the end of G_1, at the G_2–M transition, and during metaphase ([link]).



The cell cycle is controlled at three checkpoints. Integrity of the DNA is assessed at the G_1 checkpoint. Proper chromosome duplication is assessed at the G_2 checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

The G_1 Checkpoint

The G_1 checkpoint determines whether all conditions are favorable for cell division to proceed. The G_1 checkpoint, also called the restriction point, is the point at which the cell irreversibly commits to the cell-division process. In addition to adequate reserves and cell size, there is a check for damage to the genomic DNA at the G_1 checkpoint. A cell that does not meet all the requirements will not be released into the S phase.

The G_2 Checkpoint

The G_2 checkpoint bars the entry to the mitotic phase if certain conditions are not met. As in the G_1 checkpoint, cell size and protein reserves are assessed. However, the most important role of the G_2 checkpoint is to ensure that all of the chromosomes have been replicated and that the replicated DNA is not damaged.

The M Checkpoint

The M checkpoint occurs near the end of the metaphase stage of mitosis. The M checkpoint is also known as the spindle checkpoint because it determines if all the sister chromatids are correctly attached to the spindle microtubules. Because the separation of the sister chromatids during

anaphase is an irreversible step, the cycle will not proceed until the kinetochores of each pair of sister chromatids are firmly anchored to spindle fibers arising from opposite poles of the cell.

Note:
Concept in Action

Watch what occurs at the G_1, G_2, and M checkpoints by visiting this animation of the cell cycle.

Section Summary

The cell cycle is an orderly sequence of events. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages. In eukaryotes, the cell cycle consists of a long preparatory period, called interphase. Interphase is divided into G_1, S, and G_2 phases. Mitosis consists of five stages: prophase, prometaphase, metaphase, anaphase, and telophase. Mitosis is usually accompanied by cytokinesis, during which the cytoplasmic components of the daughter cells are separated either by an actin ring (animal cells) or by cell plate formation (plant cells).

Each step of the cell cycle is monitored by internal controls called checkpoints. There are three major checkpoints in the cell cycle: one near the end of G_1, a second at the G_2–M transition, and the third during metaphase.

Art Connections

Exercise:

Problem:

[link] Which of the following is the correct order of events in mitosis?

a. Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus re-forms and the cell divides. The sister chromatids separate.
b. The kinetochore becomes attached to the mitotic spindle. The sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus re-forms and the cell divides.
c. The kinetochore becomes attached to metaphase plate. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus re-forms and the cell divides.
d. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. The kinetochore breaks apart and the sister chromatids separate. The nucleus re-forms and the cell divides.

Solution:

[link] D. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. The kinetochore breaks apart and the sister chromatids separate. The nucleus reforms and the cell divides.

Multiple Choice

Exercise:

Problem:

Chromosomes are duplicated during what portion of the cell cycle?

a. G_1 phase

b. S phase
c. prophase
d. prometaphase

Solution:

B

Exercise:

Problem:

Separation of the sister chromatids is a characteristic of which stage of mitosis?

a. prometaphase
b. metaphase
c. anaphase
d. telophase

Solution:

C

Exercise:

Problem:

The individual chromosomes become visible with a light microscope during which stage of mitosis?

a. prophase
b. prometaphase
c. metaphase
d. anaphase

Solution:

A

Exercise:

Problem: What is necessary for a cell to pass the G_2 checkpoint?

a. cell has reached a sufficient size
b. an adequate stockpile of nucleotides
c. accurate and complete DNA replication
d. proper attachment of mitotic spindle fibers to kinetochores

Solution:

C

Free Response

Exercise:

Problem:

Describe the similarities and differences between the cytokinesis mechanisms found in animal cells versus those in plant cells.

Solution:

There are very few similarities between animal cell and plant cell cytokinesis. In animal cells, a ring of actin fibers is formed around the periphery of the cell at the former metaphase plate. The actin ring contracts inward, pulling the plasma membrane toward the center of the cell until the cell is pinched in two. In plant cells, a new cell wall must be formed between the daughter cells. Because of the rigid cell walls of the parent cell, contraction of the middle of the cell is not possible. Instead, a cell plate is formed in the center of the cell at the former metaphase plate. The cell plate is formed from Golgi vesicles that contain enzymes, proteins, and glucose. The vesicles fuse and the enzymes build a new cell wall from the proteins and glucose. The cell

plate grows toward, and eventually fuses with, the cell wall of the parent cell.

Glossary

anaphase
: the stage of mitosis during which sister chromatids are separated from each other

cell cycle
: the ordered sequence of events that a cell passes through between one cell division and the next

cell cycle checkpoints
: mechanisms that monitor the preparedness of a eukaryotic cell to advance through the various cell cycle stages

cell plate
: a structure formed during plant-cell cytokinesis by Golgi vesicles fusing at the metaphase plate; will ultimately lead to formation of a cell wall to separate the two daughter cells

centriole
: a paired rod-like structure constructed of microtubules at the center of each animal cell centrosome

cleavage furrow
: a constriction formed by the actin ring during animal-cell cytokinesis that leads to cytoplasmic division

cytokinesis
: the division of the cytoplasm following mitosis to form two daughter cells

G_0 phase
: a cell-cycle phase distinct from the G_1 phase of interphase; a cell in G_0 is not preparing to divide

G_1 phase

(also, first gap) a cell-cycle phase; first phase of interphase centered on cell growth during mitosis

G_2 phase

(also, second gap) a cell-cycle phase; third phase of interphase where the cell undergoes the final preparations for mitosis

interphase

the period of the cell cycle leading up to mitosis; includes G_1, S, and G_2 phases; the interim between two consecutive cell divisions

kinetochore

a protein structure in the centromere of each sister chromatid that attracts and binds spindle microtubules during prometaphase

metaphase plate

the equatorial plane midway between two poles of a cell where the chromosomes align during metaphase

metaphase

the stage of mitosis during which chromosomes are lined up at the metaphase plate

mitosis

the period of the cell cycle at which the duplicated chromosomes are separated into identical nuclei; includes prophase, prometaphase, metaphase, anaphase, and telophase

mitotic phase

the period of the cell cycle when duplicated chromosomes are distributed into two nuclei and the cytoplasmic contents are divided; includes mitosis and cytokinesis

mitotic spindle

the microtubule apparatus that orchestrates the movement of chromosomes during mitosis

prometaphase
: the stage of mitosis during which mitotic spindle fibers attach to kinetochores

prophase
: the stage of mitosis during which chromosomes condense and the mitotic spindle begins to form

quiescent
: describes a cell that is performing normal cell functions and has not initiated preparations for cell division

S phase
: the second, or synthesis phase, of interphase during which DNA replication occurs

telophase
: the stage of mitosis during which chromosomes arrive at opposite poles, decondense, and are surrounded by new nuclear envelopes

Cancer and the Cell Cycle
By the end of this section, you will be able to:

- Explain how cancer is caused by uncontrolled cell division
- Understand how proto-oncogenes are normal cell genes that, when mutated, become oncogenes
- Describe how tumor suppressors function to stop the cell cycle until certain events are completed
- Explain how mutant tumor suppressors cause cancer

Cancer is a collective name for many different diseases caused by a common mechanism: uncontrolled cell division. Despite the redundancy and overlapping levels of cell-cycle control, errors occur. One of the critical processes monitored by the cell-cycle checkpoint surveillance mechanism is the proper replication of DNA during the S phase. Even when all of the cell-cycle controls are fully functional, a small percentage of replication errors (mutations) will be passed on to the daughter cells. If one of these changes to the DNA nucleotide sequence occurs within a gene, a gene mutation results. All cancers begin when a gene mutation gives rise to a faulty protein that participates in the process of cell reproduction. The change in the cell that results from the malformed protein may be minor. Even minor mistakes, however, may allow subsequent mistakes to occur more readily. Over and over, small, uncorrected errors are passed from parent cell to daughter cells and accumulate as each generation of cells produces more non-functional proteins from uncorrected DNA damage. Eventually, the pace of the cell cycle speeds up as the effectiveness of the control and repair mechanisms decreases. Uncontrolled growth of the mutated cells outpaces the growth of normal cells in the area, and a tumor can result.

Proto-oncogenes

The genes that code for the positive cell-cycle regulators are called **proto-oncogenes**. Proto-oncogenes are normal genes that, when mutated, become **oncogenes**—genes that cause a cell to become cancerous. Consider what might happen to the cell cycle in a cell with a recently acquired oncogene. In most instances, the alteration of the DNA sequence will result in a less

functional (or non-functional) protein. The result is detrimental to the cell and will likely prevent the cell from completing the cell cycle; however, the organism is not harmed because the mutation will not be carried forward. If a cell cannot reproduce, the mutation is not propagated and the damage is minimal. Occasionally, however, a gene mutation causes a change that increases the activity of a positive regulator. For example, a mutation that allows Cdk, a protein involved in cell-cycle regulation, to be activated before it should be could push the cell cycle past a checkpoint before all of the required conditions are met. If the resulting daughter cells are too damaged to undertake further cell divisions, the mutation would not be propagated and no harm comes to the organism. However, if the atypical daughter cells are able to divide further, the subsequent generation of cells will likely accumulate even more mutations, some possibly in additional genes that regulate the cell cycle.

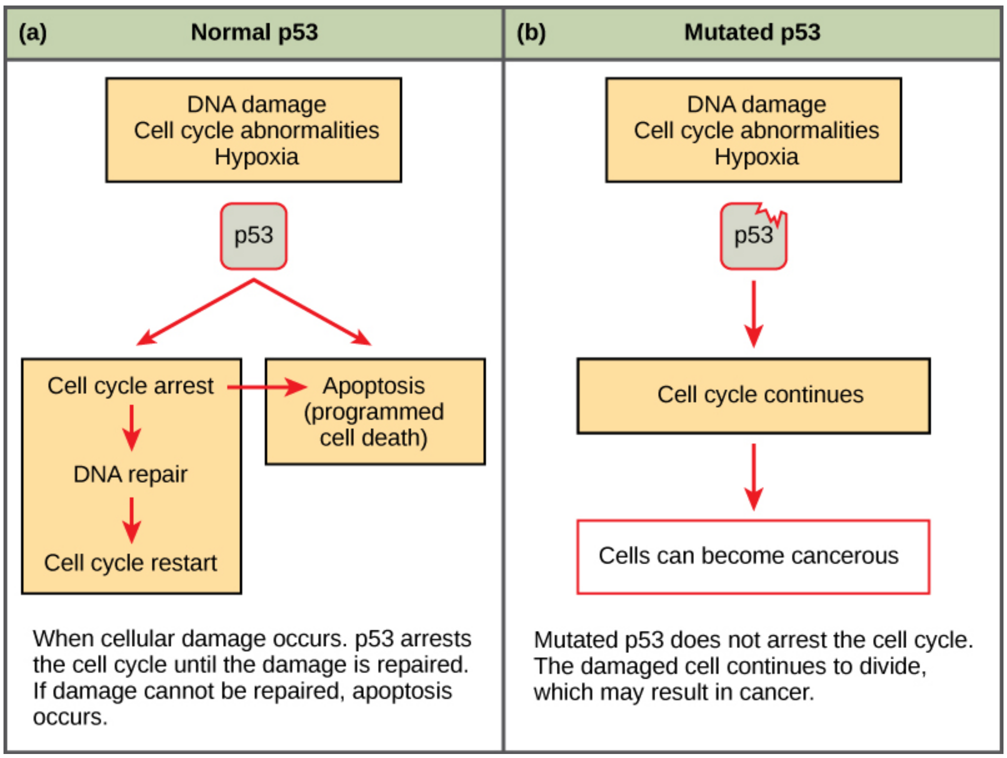
The Cdk example is only one of many genes that are considered proto-oncogenes. In addition to the cell-cycle regulatory proteins, any protein that influences the cycle can be altered in such a way as to override cell-cycle checkpoints. Once a proto-oncogene has been altered such that there is an increase in the rate of the cell cycle, it is then called an oncogene.

Tumor Suppressor Genes

Like proto-oncogenes, many of the negative cell-cycle regulatory proteins were discovered in cells that had become cancerous. **Tumor suppressor genes** are genes that code for the negative regulator proteins, the type of regulator that—when activated—can prevent the cell from undergoing uncontrolled division. The collective function of the best-understood tumor suppressor gene proteins, retinoblastoma protein (RB1), p53, and p21, is to put up a roadblock to cell-cycle progress until certain events are completed. A cell that carries a mutated form of a negative regulator might not be able to halt the cell cycle if there is a problem.

Mutated p53 genes have been identified in more than half of all human tumor cells. This discovery is not surprising in light of the multiple roles that the p53 protein plays at the G_1 checkpoint. The p53 protein activates other genes whose products halt the cell cycle (allowing time for DNA

repair), activates genes whose products participate in DNA repair, or activates genes that initiate cell death when DNA damage cannot be repaired. A damaged p53 gene can result in the cell behaving as if there are no mutations ([link]). This allows cells to divide, propagating the mutation in daughter cells and allowing the accumulation of new mutations. In addition, the damaged version of p53 found in cancer cells cannot trigger cell death.



(a) The role of p53 is to monitor DNA. If damage is detected, p53 triggers repair mechanisms. If repairs are unsuccessful, p53 signals apoptosis. (b) A cell with an abnormal p53 protein cannot repair damaged DNA and cannot signal apoptosis. Cells with abnormal p53 can become cancerous. (credit: modification of work by Thierry Soussi)

Note:
Concept in Action

Go to this website to watch an animation of how cancer results from errors in the cell cycle.

Section Summary

Cancer is the result of unchecked cell division caused by a breakdown of the mechanisms regulating the cell cycle. The loss of control begins with a change in the DNA sequence of a gene that codes for one of the regulatory molecules. Faulty instructions lead to a protein that does not function as it should. Any disruption of the monitoring system can allow other mistakes to be passed on to the daughter cells. Each successive cell division will give rise to daughter cells with even more accumulated damage. Eventually, all checkpoints become nonfunctional, and rapidly reproducing cells crowd out normal cells, resulting in tumorous growth.

Multiple Choice

Exercise:

Problem:

_________ are changes to the nucleotides in a segment of DNA that codes for a protein.

a. Proto-oncogenes
b. Tumor suppressor genes
c. Gene mutations

d. Negative regulators

Solution:

C

Exercise:

Problem:

A gene that codes for a positive cell cycle regulator is called a(n) ________.

a. kinase inhibitor
b. tumor suppressor gene
c. proto-oncogene
d. oncogene

Solution:

C

Free Response

Exercise:

Problem: Outline the steps that lead to a cell becoming cancerous.

Solution:

If one of the genes that produce regulator proteins becomes mutated, it produces a malformed, possibly non-functional, cell-cycle regulator. This increases the chance that more mutations will be left unrepaired in the cell. Each subsequent generation of cells sustains more damage. The cell cycle can speed up as a result of loss of functional checkpoint proteins. The cells can lose the ability to self-destruct.

Exercise:

Problem:

Explain the difference between a proto-oncogene and a tumor suppressor gene.

Solution:

A proto-oncogene is the segment of DNA that codes for one of the positive cell-cycle regulators. If that gene becomes mutated to a form that is overactive, it is considered an oncogene. A tumor suppressor gene is a segment of DNA that codes for one of the negative cell-cycle regulators. If that gene becomes mutated to a form that is underactive, the cell cycle will run unchecked.

Glossary

oncogene
: a mutated version of a proto-oncogene, which allows for uncontrolled progression of the cell cycle, or uncontrolled cell reproduction

proto-oncogene
: a normal gene that controls cell division by regulating the cell cycle that becomes an oncogene if it is mutated

tumor suppressor gene
: a gene that codes for regulator proteins that prevent the cell from undergoing uncontrolled division

Prokaryotic Cell Division
By the end of this section, you will be able to:

- Describe the process of binary fission in prokaryotes
- Explain how FtsZ and tubulin proteins are examples of homology

Prokaryotes such as bacteria propagate by binary fission. For unicellular organisms, cell division is the only method to produce new individuals. In both prokaryotic and eukaryotic cells, the outcome of cell reproduction is a pair of daughter cells that are genetically identical to the parent cell. In unicellular organisms, daughter cells are individuals.

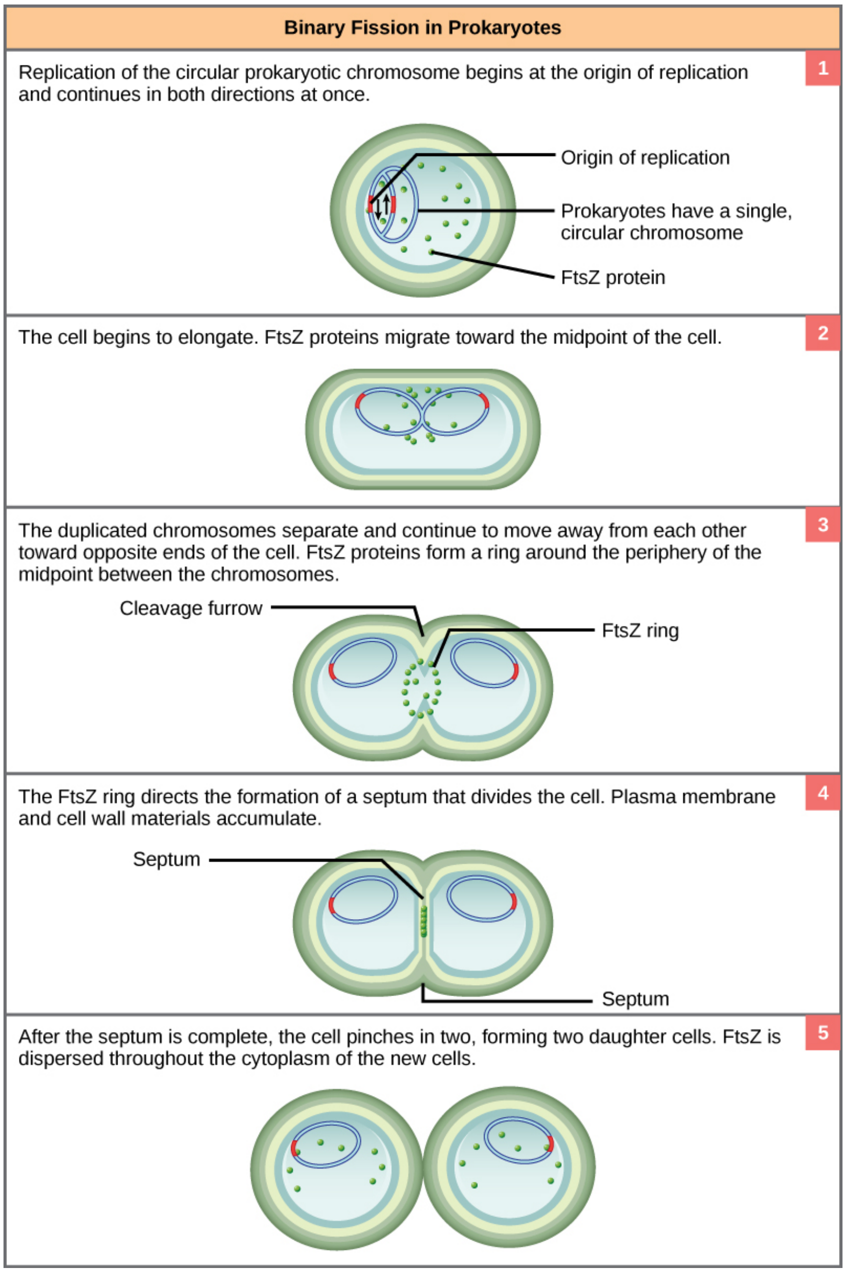
To achieve the outcome of identical daughter cells, some steps are essential. The genomic DNA must be replicated and then allocated into the daughter cells; the cytoplasmic contents must also be divided to give both new cells the machinery to sustain life. In bacterial cells, the genome consists of a single, circular DNA chromosome; therefore, the process of cell division is simplified. Mitosis is unnecessary because there is no nucleus or multiple chromosomes. This type of cell division is called binary fission.

Binary Fission

The cell division process of prokaryotes, called **binary fission**, is a less complicated and much quicker process than cell division in eukaryotes. Because of the speed of bacterial cell division, populations of bacteria can grow very rapidly. The single, circular DNA chromosome of bacteria is not enclosed in a nucleus, but instead occupies a specific location, the nucleoid, within the cell. As in eukaryotes, the DNA of the nucleoid is associated with proteins that aid in packaging the molecule into a compact size. The packing proteins of bacteria are, however, related to some of the proteins involved in the chromosome compaction of eukaryotes.

The starting point of replication, the **origin**, is close to the binding site of the chromosome to the plasma membrane ([link]). Replication of the DNA is bidirectional—moving away from the origin on both strands of the DNA loop simultaneously. As the new double strands are formed, each origin point moves away from the cell-wall attachment toward opposite ends of the cell. As the cell elongates, the growing membrane aids in the transport of the

chromosomes. After the chromosomes have cleared the midpoint of the elongated cell, cytoplasmic separation begins. A **septum** is formed between the nucleoids from the periphery toward the center of the cell. When the new cell walls are in place, the daughter cells separate.



The binary fission of a bacterium is outlined in five steps. (credit: modification of work by

“Mcstrother”/Wikimedia Commons)

Note:
Evolution in Action
Mitotic Spindle Apparatus
The precise timing and formation of the mitotic spindle is critical to the success of eukaryotic cell division. Prokaryotic cells, on the other hand, do not undergo mitosis and therefore have no need for a mitotic spindle. However, the FtsZ protein that plays such a vital role in prokaryotic cytokinesis is structurally and functionally very similar to tubulin, the building block of the microtubules that make up the mitotic spindle fibers that are necessary for eukaryotes. The formation of a ring composed of repeating units of a protein called **FtsZ** directs the partition between the nucleoids in prokaryotes. Formation of the FtsZ ring triggers the accumulation of other proteins that work together to recruit new membrane and cell-wall materials to the site. FtsZ proteins can form filaments, rings, and other three-dimensional structures resembling the way tubulin forms microtubules, centrioles, and various cytoskeleton components. In addition, both FtsZ and tubulin employ the same energy source, GTP (guanosine triphosphate), to rapidly assemble and disassemble complex structures. FtsZ and tubulin are an example of homology, structures derived from the same evolutionary origins. In this example, FtsZ is presumed to be similar to the ancestor protein to both the modern FtsZ and tubulin. While both proteins are found in extant organisms, tubulin function has evolved and diversified tremendously since the evolution from its FtsZ-like prokaryotic origin. A survey of cell-division machinery in present-day unicellular eukaryotes reveals crucial intermediary steps to the complex mitotic machinery of multicellular eukaryotes ([link]).

Mitotic Spindle Evolution			
	Structure of genetic material	**Division of nuclear material**	**Separation of daughter cells**
Prokaryotes	There is no nucleus. The single, circular chromosome exists in a region of cytoplasm called the nucleoid.	Occurs through binary fission. As the chromosome is replicated, the two copies move to opposite ends of the cell by an unknown mechanism.	FtsZ proteins assemble into a ring that pinches the cell in two.
Some protists	Linear chromosomes exist in the nucleus.	Chromosomes attach to the nuclear envelope, which remains intact. The mitotic spindle passes through the envelope and elongates the cell. No centrioles exist.	Microfilaments form a cleavage furrow that pinches the cell in two.

Mitotic Spindle Evolution			
	Structure of genetic material	**Division of nuclear material**	**Separation of daughter cells**
Other protists	Linear chromosomes exist in the nucleus.	A mitotic spindle forms from the centrioles and passes through the nuclear membrane, which remains intact. Chromosomes attach to the mitotic spindle. The mitotic spindle separates the chromosomes and elongates the cell.	Microfilaments form a cleavage furrow that pinches the cell in two.

Mitotic Spindle Evolution			
	Structure of genetic material	**Division of nuclear material**	**Separation of daughter cells**
Animal cells	Linear chromosomes exist in the nucleus.	A mitotic spindle forms from the centrioles. The nuclear envelope dissolves. Chromosomes attach to the mitotic spindle, which separates them and elongates the cell.	Microfilaments form a cleavage furrow that pinches the cell in two.

The mitotic spindle fibers of eukaryotes are composed of microtubules. Microtubules are polymers of the protein tubulin. The FtsZ protein active in prokaryote cell division is very similar to tubulin in the structures it can form and its energy source. Single-celled eukaryotes (such as yeast) display possible intermediary steps between FtsZ activity during binary fission in prokaryotes and the mitotic spindle in multicellular eukaryotes, during which the nucleus breaks down and is reformed.

Section Summary

In both prokaryotic and eukaryotic cell division, the genomic DNA is replicated and each copy is allocated into a daughter cell. The cytoplasmic contents are also divided evenly to the new cells. However, there are many differences between prokaryotic and eukaryotic cell division. Bacteria have a single, circular DNA chromosome and no nucleus. Therefore, mitosis is not necessary in bacterial cell division. Bacterial cytokinesis is directed by a ring composed of a protein called FtsZ. Ingrowth of membrane and cell-wall material from the periphery of the cells results in a septum that eventually forms the separate cell walls of the daughter cells.

Multiple Choice

Exercise:

Problem:

Which eukaryotic cell-cycle event is missing in binary fission?

a. cell growth
b. DNA duplication
c. mitosis
d. cytokinesis

Solution:

C

Exercise:

Problem:

FtsZ proteins direct the formation of a _________ that will eventually form the new cell walls of the daughter cells.

a. contractile ring
b. cell plate
c. cytoskeleton
d. septum

Solution:

D

Free Response

Exercise:

Problem:

Name the common components of eukaryotic cell division and binary fission.

Solution:

The common components of eukaryotic cell division and binary fission are DNA duplication, segregation of duplicated chromosomes, and the division of the cytoplasmic contents.

Glossary

binary fission
: the process of prokaryotic cell division

FtsZ
: a tubulin-like protein component of the prokaryotic cytoskeleton that is important in prokaryotic cytokinesis (name origin: **F**ilamenting **t**emperature-**s**ensitive mutant **Z**)

origin
: the region of the prokaryotic chromosome at which replication begins

septum
: a wall formed between bacterial daughter cells as a precursor to cell separation

Introduction
class="introduction"

Each of us, like these other large multicellular organisms, begins life as a fertilized egg. After trillions of cell divisions, each of us develops into a complex, multicellular organism. (credit a: modification of work by Frank Wouters; credit b: modification of work by Ken Cole, USGS; credit c: modification of work by Martin Pettitt)

(a) (b) (c)

The ability to reproduce *in kind* is a basic characteristic of all living things. *In kind* means that the offspring of any organism closely resembles its parent or parents. Hippopotamuses give birth to hippopotamus calves; Monterey pine trees produce seeds from which Monterey pine seedlings emerge; and adult flamingos lay eggs that hatch into flamingo chicks. *In kind* does not generally mean *exactly the same*. While many single-celled organisms and a few multicellular organisms can produce genetically identical clones of themselves through mitotic cell division, many single-celled organisms and most multicellular organisms reproduce regularly using another method.

Sexual reproduction is the production by parents of haploid cells and the fusion of a haploid cell from each parent to form a single, unique diploid cell. In multicellular organisms, the new diploid cell will then undergo mitotic cell divisions to develop into an adult organism. A type of cell division called meiosis leads to the haploid cells that are part of the sexual reproductive cycle. Sexual reproduction, specifically meiosis and fertilization, introduces variation into offspring that may account for the evolutionary success of sexual reproduction. The vast majority of eukaryotic organisms can or must employ some form of meiosis and fertilization to reproduce.

Sexual Reproduction
By the end of this section, you will be able to:

- Explain that variation among offspring is a potential evolutionary advantage resulting from sexual reproduction
- Describe the three different life-cycle strategies among sexual multicellular organisms and their commonalities

Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. The fact that most eukaryotes reproduce sexually is evidence of its evolutionary success. In many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction. On the surface, offspring that are genetically identical to the parent may appear to be more advantageous. If the parent organism is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit to an organism that can produce offspring by asexual budding, fragmentation, or asexual eggs. These methods of reproduction do not require another organism of the opposite sex. There is no need to expend energy finding or attracting a mate. That energy can be spent on producing more offspring. Indeed, some organisms that lead a solitary lifestyle have retained the ability to reproduce asexually. In addition, asexual populations only have female individuals, so every individual is capable of reproduction. In contrast, the males in sexual populations (half the population) are not producing offspring themselves. Because of this, an asexual population can grow twice as fast as a sexual population in theory. This means that in competition, the asexual population would have the advantage. All of these advantages to asexual reproduction, which are also disadvantages to sexual reproduction, should mean that the number of species with asexual reproduction should be more common.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why is sexual reproduction so common? This is one of the important questions in biology and has been the focus of much research from the latter half of the twentieth century until now. A likely explanation is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of those

offspring. The only source of variation in asexual organisms is mutation. This is the ultimate source of variation in sexual organisms. In addition, those different mutations are continually reshuffled from one generation to the next when different parents combine their unique genomes, and the genes are mixed into different combinations by the process of **meiosis**. Meiosis is the division of the contents of the nucleus that divides the chromosomes among gametes. Variation is introduced during meiosis, as well as when the gametes combine in fertilization.

Note:
Evolution in Action

The Red Queen Hypothesis

There is no question that sexual reproduction provides evolutionary advantages to organisms that employ this mechanism to produce offspring. The problematic question is why, even in the face of fairly stable conditions, sexual reproduction persists when it is more difficult and produces fewer offspring for individual organisms? Variation is the outcome of sexual reproduction, but why are ongoing variations necessary? Enter the Red Queen hypothesis, first proposed by Leigh Van Valen in 1973.[footnote] The concept was named in reference to the Red Queen's race in Lewis Carroll's book, *Through the Looking-Glass*, in which the Red Queen says one must run at full speed just to stay where one is.

Leigh Van Valen, “A new evolutionary law,” Evolutionary Theory 1 (1973): 1–30.

All species coevolve with other organisms. For example, predators coevolve with their prey, and parasites coevolve with their hosts. A remarkable example of coevolution between predators and their prey is the unique coadaptation of night flying bats and their moth prey. Bats find their prey by emitting high-pitched clicks, but moths have evolved simple ears to hear these clicks so they can avoid the bats. The moths have also adapted behaviors, such as flying away from the bat when they first hear it, or dropping suddenly to the ground when the bat is upon them. Bats have evolved “quiet” clicks in an attempt to evade the moth’s hearing. Some moths have evolved the ability to respond to the bats’ clicks with their own clicks as a strategy to confuse the bats echolocation abilities.

Each tiny advantage gained by favorable variation gives a species an edge over close competitors, predators, parasites, or even prey. The only method that will allow a coevolving species to keep its own share of the resources is also to continually improve its ability to survive and produce offspring. As one species gains an advantage, other species must also develop an advantage or they will be outcompeted. No single species progresses too far ahead because genetic variation among progeny of sexual reproduction provides all species with a mechanism to produce adapted individuals. Species whose individuals cannot keep up become extinct. The Red Queen’s catchphrase was, “It takes all the running you can do to stay in the same place.” This is an apt description of coevolution between competing species.

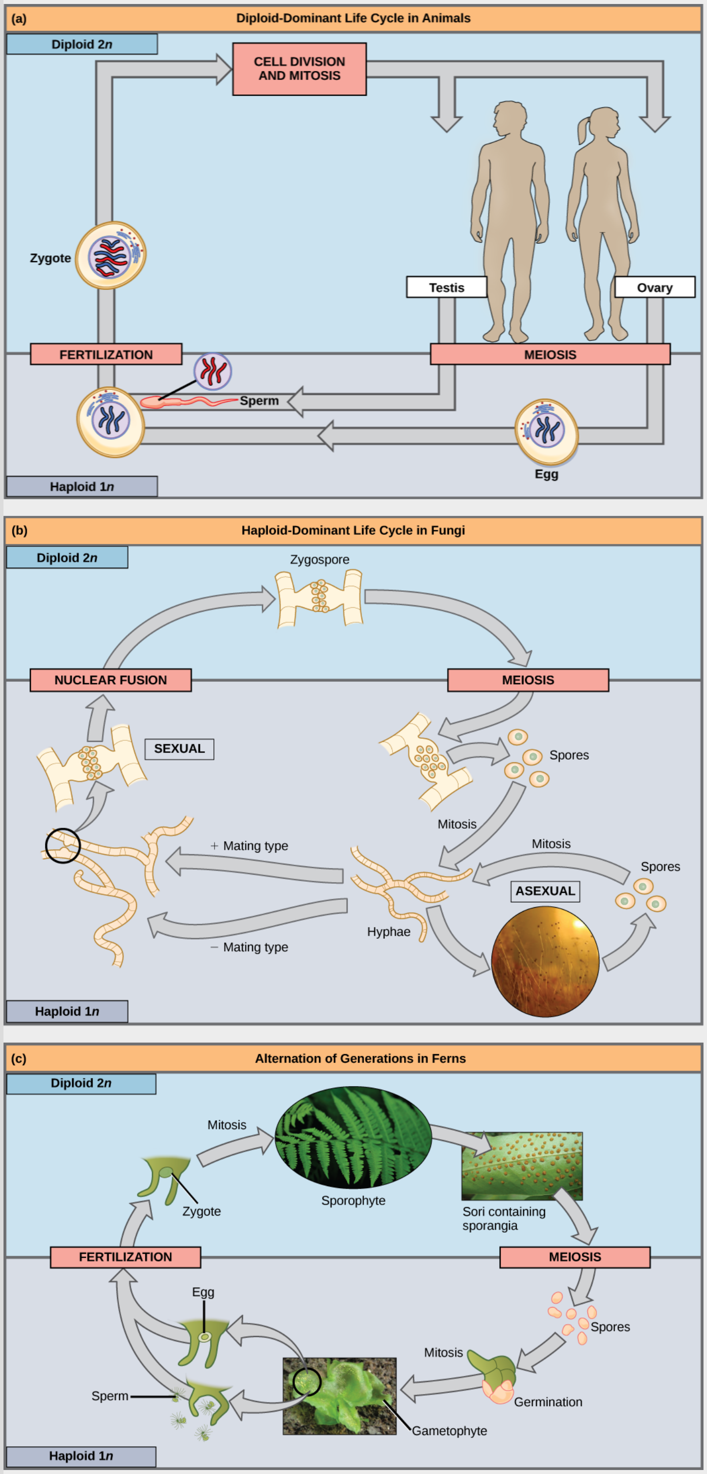
Life Cycles of Sexually Reproducing Organisms

Fertilization and meiosis alternate in sexual **life cycles**. What happens between these two events depends on the organism. The process of meiosis reduces the resulting gamete’s chromosome number by half. Fertilization, the joining of two haploid gametes, restores the diploid condition. There are three main categories of life cycles in multicellular organisms: **diploid-dominant**, in which the multicellular diploid stage is the most obvious life stage (and there is no multicellular haploid stage), as with most animals including humans; **haploid-dominant**, in which the multicellular haploid stage is the most obvious life stage (and there is no multicellular diploid stage), as with all fungi and some algae; and **alternation of generations**, in which the two stages, haploid and diploid, are apparent to one degree or another depending on the group, as with plants and some algae.

Nearly all animals employ a diploid-dominant life-cycle strategy in which the only haploid cells produced by the organism are the gametes. The gametes are produced from diploid **germ cells**, a special cell line that only produces gametes. Once the haploid gametes are formed, they lose the ability to divide again. There is no multicellular haploid life stage. Fertilization occurs with the fusion of two gametes, usually from different individuals, restoring the diploid state ([link]**a**).

Note:

Art Connection



(a) In animals, sexually reproducing adults form haploid gametes from

diploid germ cells. (b) Fungi, such as black bread mold (*Rhizopus nigricans*), have haploid-dominant life cycles. (c) Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. (credit c "fern": modification of work by Cory Zanker; credit c "gametophyte": modification of work by "Vlmastra"/Wikimedia Commons)

If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

Most fungi and algae employ a life-cycle strategy in which the multicellular "body" of the organism is haploid. During sexual reproduction, specialized haploid cells from two individuals join to form a diploid zygote. The zygote immediately undergoes meiosis to form four haploid cells called spores ([link]**b**).

The third life-cycle type, employed by some algae and all plants, is called alternation of generations. These species have both haploid and diploid multicellular organisms as part of their life cycle. The haploid multicellular plants are called **gametophytes** because they produce gametes. Meiosis is not involved in the production of gametes in this case, as the organism that produces gametes is already haploid. Fertilization between the gametes forms a diploid zygote. The zygote will undergo many rounds of mitosis and give rise to a diploid multicellular plant called a **sporophyte**. Specialized cells of the sporophyte will undergo meiosis and produce haploid spores. The spores will develop into the gametophytes ([link]**c**).

Section Summary

Nearly all eukaryotes undergo sexual reproduction. The variation introduced into the reproductive cells by meiosis appears to be one of the advantages of sexual reproduction that has made it so successful. Meiosis and fertilization alternate in sexual life cycles. The process of meiosis produces genetically unique reproductive cells called gametes, which have half the number of chromosomes as the parent cell. Fertilization, the fusion of haploid gametes from two individuals, restores the diploid condition. Thus, sexually reproducing organisms alternate between haploid and diploid stages. However, the ways in which reproductive cells are produced and the timing between meiosis and fertilization vary greatly. There are three main categories of life cycles: diploid-dominant, demonstrated by most animals; haploid-dominant, demonstrated by all fungi and some algae; and alternation of generations, demonstrated by plants and some algae.

Art Connections

Exercise:

Problem:

[link] If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

Solution:

[link] Yes, it will be able to reproduce asexually.

Multiple Choice

Exercise:

Problem:

What is a likely evolutionary advantage of sexual reproduction over asexual reproduction?

a. sexual reproduction involves fewer steps
b. less chance of using up the resources in a given environment

c. sexual reproduction results in greater variation in the offspring
d. sexual reproduction is more cost-effective

Solution:

C

Exercise:

Problem:

Which type of life cycle has both a haploid and diploid multicellular stage?

a. an asexual life cycle
b. diploid-dominant
c. haploid-dominant
d. alternation of generations

Solution:

D

Exercise:

Problem: Which event leads to a diploid cell in a life cycle?

a. meiosis
b. fertilization
c. alternation of generations
d. mutation

Solution:

B

Free Response

Exercise:

Problem:

Explain the advantage that populations of sexually reproducing organisms have over asexually reproducing organisms?

Solution:

The offspring of sexually reproducing organisms are all genetically unique. Because of this, sexually reproducing organisms may have more successful survival of offspring in environments that change than asexually reproducing organisms, whose offspring are all genetically identical. In addition, the rate of adaptation of sexually reproducing organisms is higher, because of their increased variation. This may allow sexually reproducing organisms to adapt more quickly to competitors and parasites, who are evolving new ways to exploit or outcompete them.

Exercise:

Problem:

Describe the two events that are common to all sexually reproducing organisms and how they fit into the different life cycles of those organisms.

Solution:

The two events common to all sexually reproducing organisms are meiosis and fertilization. Meiosis reduces a diploid cell to a haploid state. The haploid cell may divide mitotically to produce an organism, some of whose cells will combine during fertilization, or the haploid cells produced by meiosis may immediately combine in fertilization to produce a diploid cell that divides to produce an organism.

Glossary

alternation of generations
: a life-cycle type in which the diploid and haploid stages alternate

diploid-dominant
: a life-cycle type in which the multicellular diploid stage is prevalent

haploid-dominant
: a life-cycle type in which the multicellular haploid stage is prevalent

gametophyte
: a multicellular haploid life-cycle stage that produces gametes

germ cell
: a specialized cell that produces gametes, such as eggs or sperm

life cycle
: the sequence of events in the development of an organism and the production of cells that produce offspring

meiosis
: a nuclear division process that results in four haploid cells

sporophyte
: a multicellular diploid life-cycle stage that produces spores

Meiosis

By the end of this section, you will be able to:

- Describe the behavior of chromosomes during meiosis
- Describe cellular events during meiosis
- Explain the differences between meiosis and mitosis
- Explain the mechanisms within meiosis that generate genetic variation among the products of meiosis

Sexual reproduction requires **fertilization**, a union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. The number of sets of chromosomes in a cell is called its ploidy level. Haploid cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid. If the reproductive cycle is to continue, the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division, known as meiosis, that reduces the number of chromosome sets.

Most animals and plants are diploid, containing two sets of chromosomes; in each **somatic cell** (the nonreproductive cells of a multicellular organism), the nucleus contains two copies of each chromosome that are referred to as homologous chromosomes. Somatic cells are sometimes referred to as “body” cells. Homologous chromosomes are matched pairs containing genes for the same traits in identical locations along their length. Diploid organisms inherit one copy of each homologous chromosome from each parent; all together, they are considered a full set of chromosomes. In animals, haploid cells containing a single copy of each homologous chromosome are found only within gametes. Gametes fuse with another haploid gamete to produce a diploid cell.

The nuclear division that forms haploid cells, which is called meiosis, is related to mitosis. As you have learned, mitosis is part of a cell reproduction cycle that results in identical daughter nuclei that are also genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei contain the same number of chromosome sets—diploid for

most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve the reduction in chromosome number, meiosis consists of one round of chromosome duplication and two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the stages are designated with a “I” or “II.” Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, prometaphase I, and so on. Meiosis I reduces the number of chromosome sets from two to one. The genetic information is also mixed during this division to create unique recombinant chromosomes. **Meiosis II**, in which the second round of meiotic division takes place in a way that is similar to mitosis, includes prophase II, prometaphase II, and so on.

Interphase

Meiosis is preceded by an interphase consisting of the G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase is the first phase of interphase and is focused on cell growth. In the S phase, the DNA of the chromosomes is replicated. Finally, in the G_2 phase, the cell undergoes the final preparations for meiosis.

During DNA duplication of the S phase, each chromosome becomes composed of two identical copies (called sister chromatids) that are held together at the centromere until they are pulled apart during meiosis II. In an animal cell, the centrosomes that organize the microtubules of the meiotic spindle also replicate. This prepares the cell for the first meiotic phase.

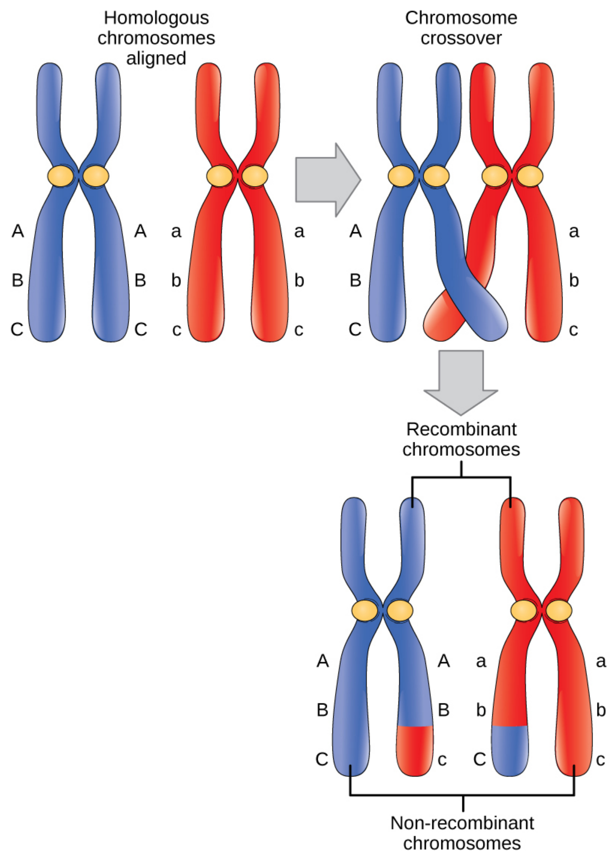
Meiosis I

Early in prophase I, the chromosomes can be seen clearly microscopically. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. The tight pairing of the homologous chromosomes is called **synapsis**. In synapsis, the genes on the chromatids of the homologous chromosomes are precisely

aligned with each other. An exchange of chromosome segments between non-sister homologous chromatids occurs and is called **crossing over**. This process is revealed visually after the exchange as **chiasmata** (singular = *chiasma*) ([link]).

As prophase I progresses, the close association between homologous chromosomes begins to break down, and the chromosomes continue to condense, although the homologous chromosomes remain attached to each other at chiasmata. The number of chiasmata varies with the species and the length of the chromosome. At the end of prophase I, the pairs are held together only at chiasmata ([link]) and are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

The crossover events are the first source of genetic variation produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete, it will carry some DNA from one parent of the individual and some DNA from the other parent. The **recombinant** sister chromatid has a combination of maternal and paternal genes that did not exist before the crossover.



In this illustration of the effects of crossing over, the blue chromosome came from the individual's father and the red chromosome came from the individual's mother. Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes. The chromosomes that have a mixture of maternal and paternal sequence are called recombinant and the chromosomes that are

completely paternal or maternal
are called non-recombinant.

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. The microtubules assembled from centrosomes at opposite poles of the cell grow toward the middle of the cell. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome attached at one pole and the other homologous chromosome attached to the other pole. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

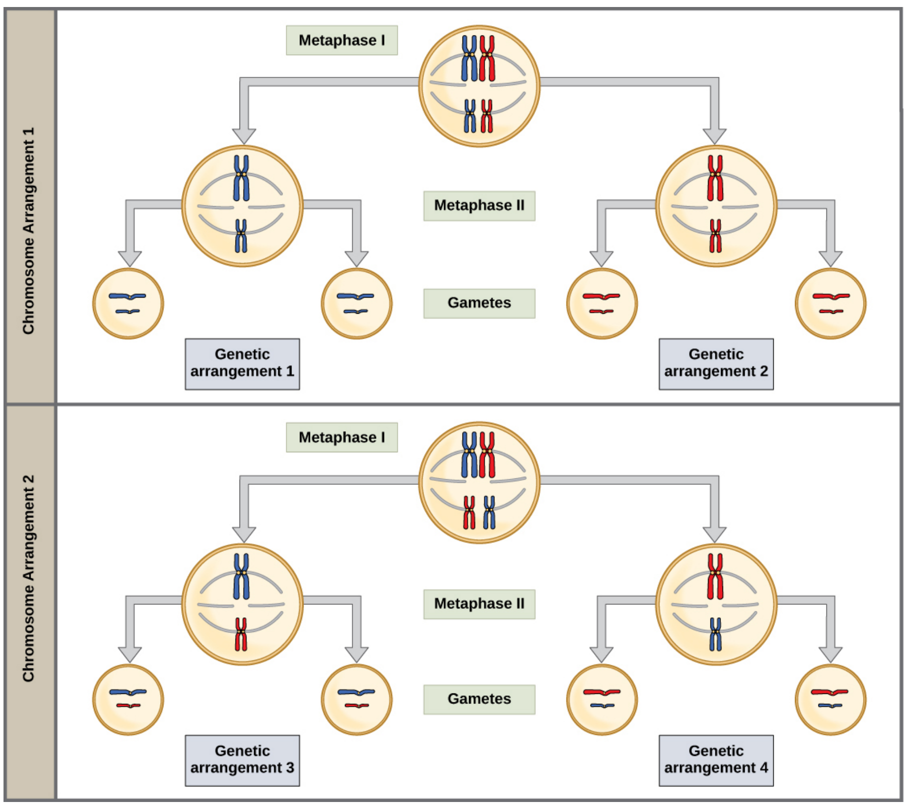
During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The orientation of each pair of homologous chromosomes at the center of the cell is random.

This randomness, called independent assortment, is the physical basis for the generation of the second form of genetic variation in offspring. Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. In metaphase I, these pairs line up at the midway point between the two poles of the cell. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations depends on the number of chromosomes making up a set. There are two possibilities for orientation (for each tetrad); thus, the possible number of alignments equals 2^n where n is the number of

chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possibilities. This number does not include the variability previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition ([link]).

To summarize the genetic consequences of meiosis I: the maternal and paternal genes are recombined by crossover events occurring on each homologous pair during prophase I; in addition, the random assortment of tetrads at metaphase produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.



To demonstrate random, independent assortment at metaphase I, consider a cell with $n = 2$. In this case, there are two possible arrangements at the equatorial plane in metaphase I, as shown in the upper cell of each panel. These two possible

orientations lead to the production of genetically different gametes. With more chromosomes, the number of possible arrangements increases dramatically.

In anaphase I, the spindle fibers pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. It is the chiasma connections that are broken in anaphase I as the fibers attached to the fused kinetochores pull the homologous chromosomes apart ([link]).

In telophase I, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the chromatids in telophase I.

Cytokinesis, the physical separation of the cytoplasmic components into two daughter cells, occurs without reformation of the nuclei in other organisms. In nearly all species, cytokinesis separates the cell contents by either a cleavage furrow (in animals and some fungi), or a cell plate that will ultimately lead to formation of cell walls that separate the two daughter cells (in plants). At each pole, there is just one member of each pair of the homologous chromosomes, so only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though there are duplicate copies of the set because each homolog still consists of two sister chromatids that are still attached to each other. However, although the sister chromatids were once duplicates of the same chromosome, they are no longer identical at this stage because of crossovers.

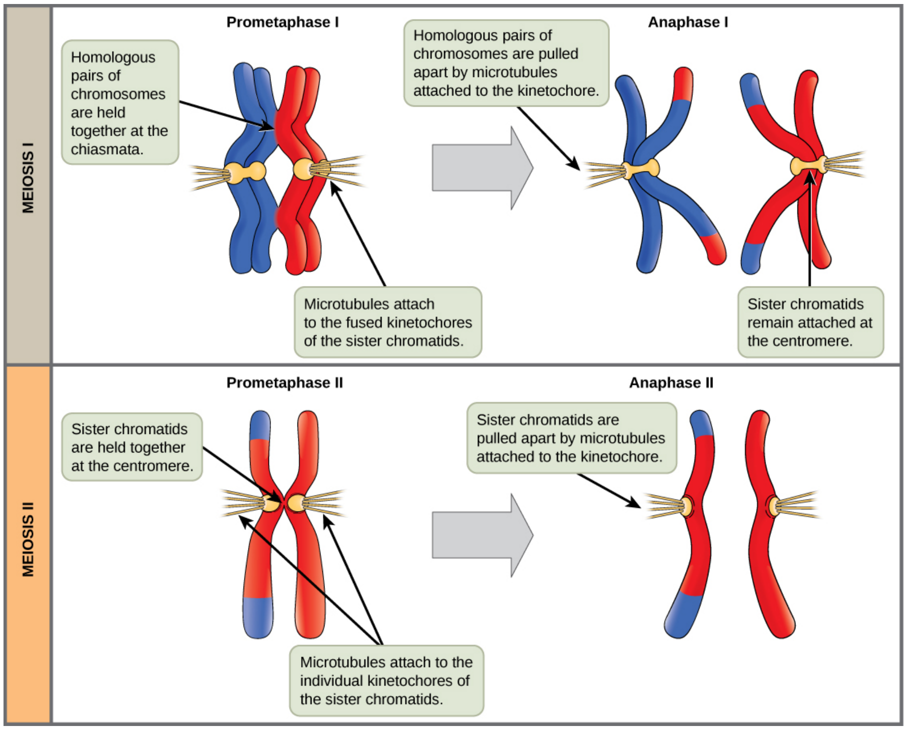
Note:
Concept in Action

Review the process of meiosis, observing how chromosomes align and migrate, at this site.

Meiosis II

In meiosis II, the connected sister chromatids remaining in the haploid cells from meiosis I will be split to form four haploid cells. In some species, cells enter a brief interphase, or **interkinesis**, that lacks an S phase, before entering meiosis II. Chromosomes are not duplicated during interkinesis. The two cells produced in meiosis I go through the events of meiosis II in synchrony. Overall, meiosis II resembles the mitotic division of a haploid cell.

In prophase II, if the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed. In prometaphase II, the nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles. In metaphase II, the sister chromatids are maximally condensed and aligned at the center of the cell. In anaphase II, the sister chromatids are pulled apart by the spindle fibers and move toward opposite poles.



In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to individual kinetochores of sister chromatids. In anaphase II, the sister chromatids are separated.

In telophase II, the chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four genetically unique haploid cells. At this point, the nuclei in the newly produced cells are both haploid and have only one copy of the single set of chromosomes. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombination of maternal and paternal segments of chromosomes—with their sets of genes—that occurs during crossover.

Comparing Meiosis and Mitosis

Mitosis and meiosis, which are both forms of division of the nucleus in eukaryotic cells, share some similarities, but also exhibit distinct differences that lead to their very different outcomes. Mitosis is a single nuclear division that results in two nuclei, usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original. They have the same number of sets of chromosomes: one in the case of haploid cells, and two in the case of diploid cells. On the other hand, meiosis is two nuclear divisions that result in four nuclei, usually partitioned into four new cells. The nuclei resulting from meiosis are never genetically identical, and they contain one chromosome set only—this is half the number of the original cell, which was diploid ([link]).

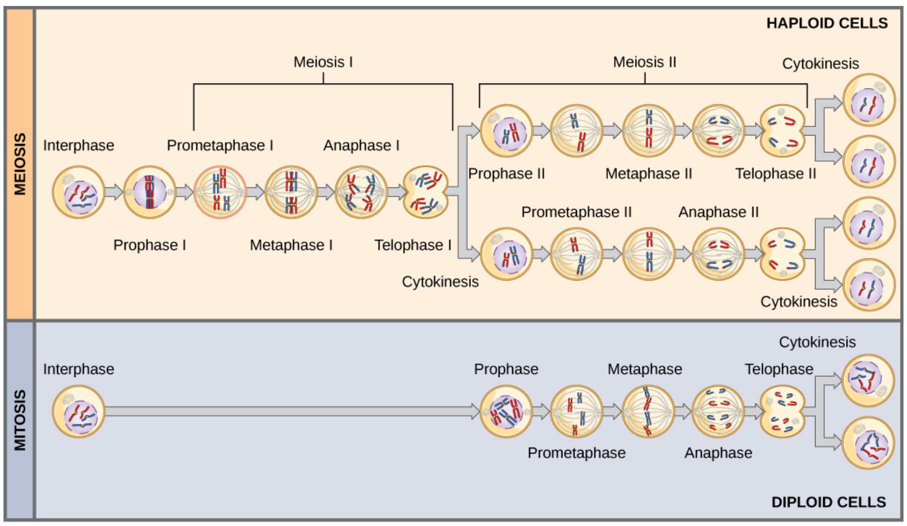
The differences in the outcomes of meiosis and mitosis occur because of differences in the behavior of the chromosomes during each process. Most of these differences in the processes occur in meiosis I, which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together, experience chiasmata and crossover between sister chromatids, and line up along the metaphase plate in tetrads with spindle fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I, never in mitosis.

Homologous chromosomes move to opposite poles during meiosis I so the number of sets of chromosomes in each nucleus-to-be is reduced from two to one. For this reason, meiosis I is referred to as a **reduction division**. There is no such reduction in ploidy level in mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, duplicated chromosomes (only one set of them) line up at the center of the cell with divided kinetochores attached to spindle fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid is pulled to one pole and the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossovers, the two products of each meiosis II division would be identical as in mitosis; instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because, although there are fewer copies of the genome in the resulting

cells, there is still one set of chromosomes, as there was at the end of meiosis I.

Cells produced by mitosis will function in different parts of the body as a part of growth or replacing dead or damaged cells. They may even be involved in asexual reproduction in some organisms. Cells produced by meiosis in a diploid-dominant organism such as an animal will only participate in sexual reproduction.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

Note:
Concept in Action

For an animation comparing mitosis and meiosis, go to this website.

Section Summary

Sexual reproduction requires that diploid organisms produce haploid cells that can fuse during fertilization to form diploid offspring. The process that results in haploid cells is called meiosis. Meiosis is a series of events that arrange and separate chromosomes into daughter cells. During the interphase of meiosis, each chromosome is duplicated. In meiosis, there are two rounds of nuclear division resulting in four nuclei and usually four haploid daughter cells, each with half the number of chromosomes as the parent cell. During meiosis, variation in the daughter nuclei is introduced because of crossover in prophase I and random alignment at metaphase I. The cells that are produced by meiosis are genetically unique.

Meiosis and mitosis share similarities, but have distinct outcomes. Mitotic divisions are single nuclear divisions that produce daughter nuclei that are genetically identical and have the same number of chromosome sets as the original cell. Meiotic divisions are two nuclear divisions that produce four daughter nuclei that are genetically different and have one chromosome set rather than the two sets the parent cell had. The main differences between the processes occur in the first division of meiosis. The homologous

chromosomes separate into different nuclei during meiosis I causing a reduction of ploidy level. The second division of meiosis is much more similar to a mitotic division.

Multiple Choice

Exercise:

Problem: Meiosis produces _________ daughter cells.

a. two haploid
b. two diploid
c. four haploid
d. four diploid

Solution:

C

Exercise:

Problem:

At which stage of meiosis are sister chromatids separated from each other?

a. prophase I
b. prophase II
c. anaphase I
d. anaphase II

Solution:

D

Exercise:

Problem: The part of meiosis that is similar to mitosis is ________.

a. meiosis I
b. anaphase I
c. meiosis II
d. interkinesis

Solution:

C

Exercise:

Problem:

If a muscle cell of a typical organism has 32 chromosomes, how many chromosomes will be in a gamete of that same organism?

a. 8
b. 16
c. 32
d. 64

Solution:

B

Free Response

Exercise:

Problem:

Explain how the random alignment of homologous chromosomes during metaphase I contributes to variation in gametes produced by meiosis.

Solution:

Random alignment leads to new combinations of traits. The chromosomes that were originally inherited by the gamete-producing individual came equally from the egg and the sperm. In metaphase I, the duplicated copies of these maternal and paternal homologous chromosomes line up across the center of the cell to form a tetrad. The orientation of each tetrad is random. There is an equal chance that the maternally derived chromosomes will be facing either pole. The same is true of the paternally derived chromosomes. The alignment should occur differently in almost every meiosis. As the homologous chromosomes are pulled apart in anaphase I, any combination of maternal and paternal chromosomes will move toward each pole. The gametes formed from these two groups of chromosomes will have a mixture of traits from the individual's parents. Each gamete is unique.

Exercise:

Problem:

In what ways is meiosis II similar to and different from mitosis of a diploid cell?

Solution:

The two divisions are similar in that the chromosomes line up along the metaphase plate individually, meaning unpaired with other chromosomes (as in meiosis I). In addition, each chromosome consists of two sister chromatids that will be pulled apart. The two divisions are different because in meiosis II there are half the number of chromosomes that are present in a diploid cell of the same species undergoing mitosis. This is because meiosis I reduced the number of chromosomes to a haploid state.

Glossary

chiasmata

(singular = *chiasma*) the structure that forms at the crossover points after genetic material is exchanged

crossing over
: (also, recombination) the exchange of genetic material between homologous chromosomes resulting in chromosomes that incorporate genes from both parents of the organism forming reproductive cells

fertilization
: the union of two haploid cells typically from two individual organisms

interkinesis
: a period of rest that may occur between meiosis I and meiosis II; there is no replication of DNA during interkinesis

meiosis I
: the first round of meiotic cell division; referred to as reduction division because the resulting cells are haploid

meiosis II
: the second round of meiotic cell division following meiosis I; sister chromatids are separated from each other, and the result is four unique haploid cells

recombinant
: describing something composed of genetic material from two sources, such as a chromosome with both maternal and paternal segments of DNA

reduction division
: a nuclear division that produces daughter nuclei each having one-half as many chromosome sets as the parental nucleus; meiosis I is a reduction division

somatic cell
: all the cells of a multicellular organism except the gamete-forming cells

synapsis
the formation of a close association between homologous chromosomes during prophase I

tetrad
two duplicated homologous chromosomes (four chromatids) bound together by chiasmata during prophase I

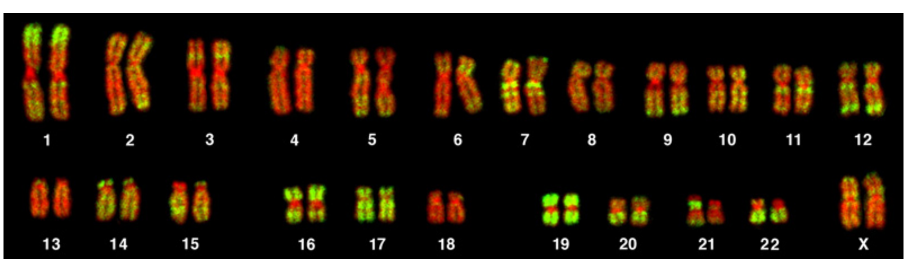
Errors in Meiosis
By the end of this section, you will be able to:

- Explain how nondisjunction leads to disorders in chromosome number
- Describe how errors in chromosome structure occur through inversions and translocations

Inherited disorders can arise when chromosomes behave abnormally during meiosis. Chromosome disorders can be divided into two categories: abnormalities in chromosome number and chromosome structural rearrangements. Because even small segments of chromosomes can span many genes, chromosomal disorders are characteristically dramatic and often fatal.

Disorders in Chromosome Number

The isolation and microscopic observation of chromosomes forms the basis of cytogenetics and is the primary method by which clinicians detect chromosomal abnormalities in humans. A **karyotype** is the number and appearance of chromosomes, including their length, banding pattern, and centromere position. To obtain a view of an individual's karyotype, cytologists photograph the chromosomes and then cut and paste each chromosome into a chart, or **karyogram** ([link]).



This karyogram shows the chromosomes of a female human immune cell during mitosis. (credit: Andreas Bolzer, et al)

Note:

Careers in Action

Geneticists Use Karyograms to Identify Chromosomal Aberrations

The karyotype is a method by which traits characterized by chromosomal abnormalities can be identified from a single cell. To observe an individual’s karyotype, a person’s cells (like white blood cells) are first collected from a blood sample or other tissue. In the laboratory, the isolated cells are stimulated to begin actively dividing. A chemical is then applied to the cells to arrest mitosis during metaphase. The cells are then fixed to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize the distinct and reproducible banding patterns of each chromosome pair. Following staining, chromosomes are viewed using bright-field microscopy. An experienced cytogeneticist can identify each band. In addition to the banding patterns, chromosomes are further identified on the basis of size and centromere location. To obtain the classic depiction of the karyotype in which homologous pairs of chromosomes are aligned in numerical order from longest to shortest, the geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into this pattern ([link]).

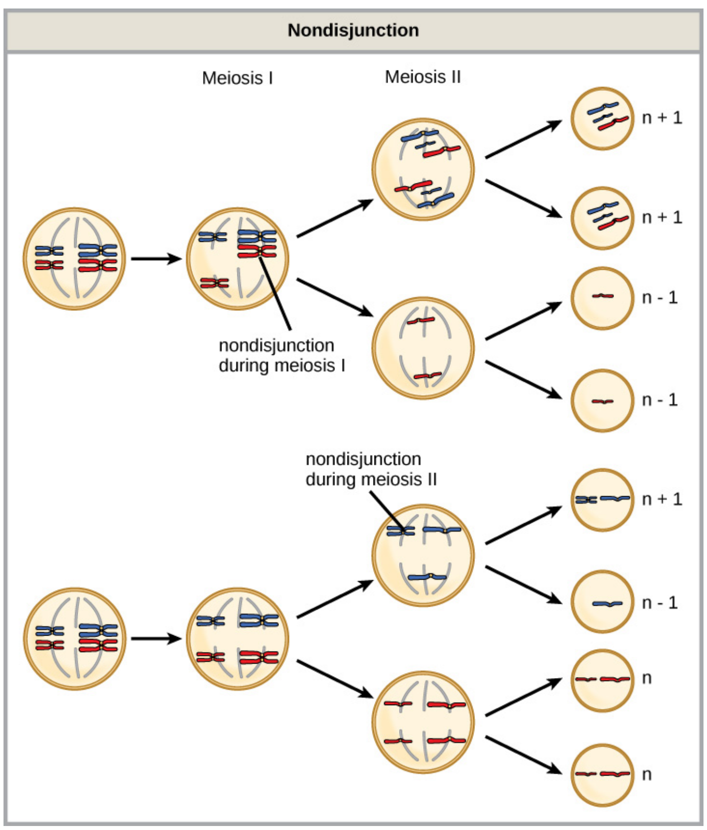
At its most basic, the karyogram may reveal genetic abnormalities in which an individual has too many or too few chromosomes per cell. Examples of this are Down syndrome, which is identified by a third copy of chromosome 21, and Turner syndrome, which is characterized by the presence of only one X chromosome in women instead of two. Geneticists can also identify large deletions or insertions of DNA. For instance, Jacobsen syndrome, which involves distinctive facial features as well as heart and bleeding defects, is identified by a deletion on chromosome 11. Finally, the karyotype can pinpoint **translocations**, which occur when a segment of genetic material breaks from one chromosome and reattaches to another chromosome or to a different part of the same chromosome. Translocations are implicated in certain cancers, including chronic myelogenous leukemia.

By observing a karyogram, geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring even before birth.

Nondisjunctions, Duplications, and Deletions

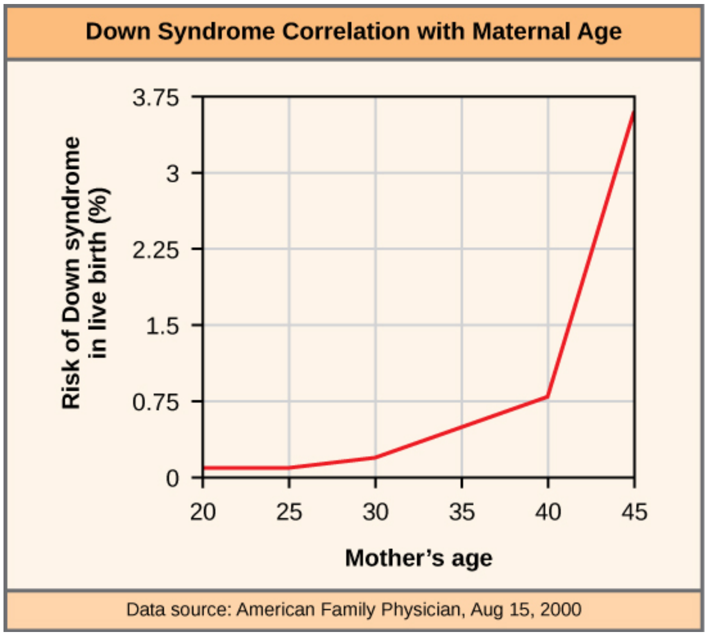
Of all the chromosomal disorders, abnormalities in chromosome number are the most easily identifiable from a karyogram. Disorders of chromosome number include the duplication or loss of entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when pairs of homologous chromosomes or sister chromatids fail to separate during meiosis. The risk of nondisjunction increases with the age of the parents.

Nondisjunction can occur during either meiosis I or II, with different results ([link]). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that chromosome and two gametes with two copies of the chromosome. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one copy of the chromosome, and one gamete with two copies of the chromosome.



Following meiosis, each gamete has one copy of each chromosome. Nondisjunction occurs when homologous chromosomes (meiosis I) or sister chromatids (meiosis II) fail to separate during meiosis.

An individual with the appropriate number of chromosomes for their species is called **euploid**; in humans, euploidy corresponds to 22 pairs of **autosomes** and one pair of sex chromosomes. An individual with an error in chromosome number is described as **aneuploid**, a term that includes **monosomy** (loss of one chromosome) or **trisomy** (gain of an extraneous chromosome). Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they have only one copy of essential genes. Most autosomal trisomies also fail to develop to birth; however, duplications of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from a different type of genetic imbalance: an excess in gene dose. Cell functions are calibrated to the amount of gene product produced by two copies (doses) of each gene; adding a third copy (dose) disrupts this balance. The most common trisomy is that of chromosome 21, which leads to Down syndrome. Individuals with this inherited disorder have characteristic physical features and developmental delays in growth and cognition. The incidence of Down syndrome is correlated with maternal age, such that older women are more likely to give birth to children with Down syndrome ([link]).



The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

Note:
Concept in Action

Visualize the addition of a chromosome that leads to Down syndrome [in this video](#).

Humans display dramatic deleterious effects with autosomal trisomies and monosomies. Therefore, it may seem counterintuitive that human females and males can function normally, despite carrying different numbers of the X chromosome. In part, this occurs because of a process called **X inactivation**. Early in development, when female mammalian embryos consist of just a few thousand cells, one X chromosome in each cell inactivates by condensing into a structure called a Barr body. The genes on the inactive X chromosome are not expressed. The particular X chromosome (maternally or paternally derived) that is inactivated in each cell is random, but once the inactivation occurs, all cells descended from that cell will have the same inactive X chromosome. By this process, females compensate for their double genetic dose of X chromosome.

In so-called "tortoiseshell" cats, X inactivation is observed as coat-color variegation ([link]). Females heterozygous for an X-linked coat color gene will express one of two different coat colors over different regions of their body, corresponding to whichever X chromosome is inactivated in the embryonic cell progenitor of that region. When you see a tortoiseshell cat, you will know that it has to be a female.

Embryonic inactivation of one of two different X

chromosomes encoding different coat colors gives rise to the tortoiseshell phenotype in cats. (credit: Michael Bodega)

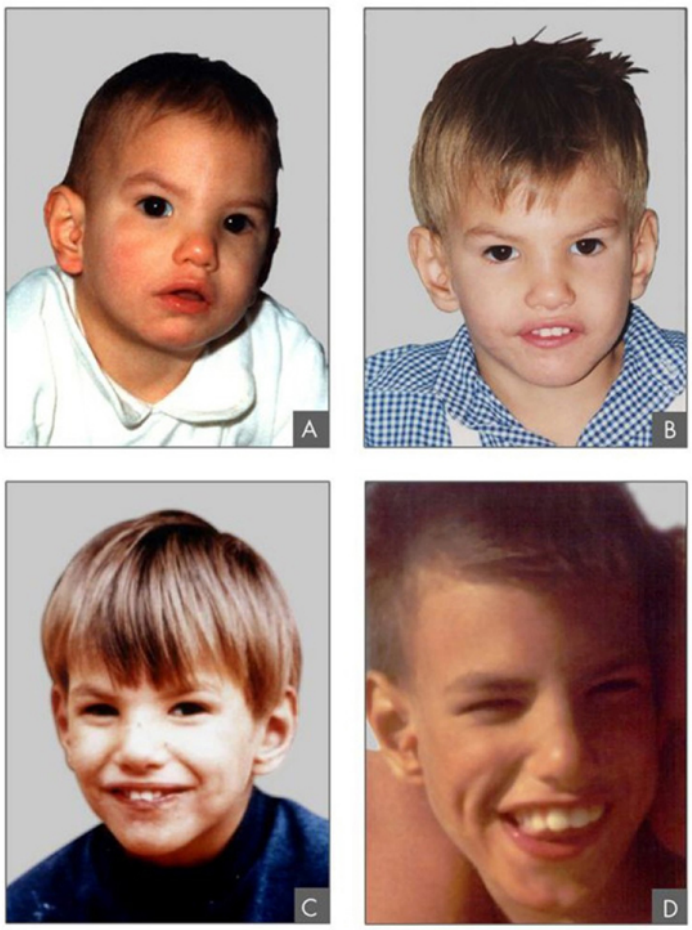
In an individual carrying an abnormal number of X chromosomes, cellular mechanisms will inactivate all but one X in each of her cells. As a result, X-chromosomal abnormalities are typically associated with mild mental and physical defects, as well as sterility. If the X chromosome is absent altogether, the individual will not develop.

Several errors in sex chromosome number have been characterized. Individuals with three X chromosomes, called triplo-X, appear female but express developmental delays and reduced fertility. The XXY chromosome complement, corresponding to one type of Klinefelter syndrome, corresponds to male individuals with small testes, enlarged breasts, and reduced body hair. The extra X chromosome undergoes inactivation to compensate for the excess genetic dosage. Turner syndrome, characterized as an X0 chromosome complement (i.e., only a single sex chromosome), corresponds to a female individual with short stature, webbed skin in the neck region, hearing and cardiac impairments, and sterility.

An individual with more than the correct number of chromosome sets (two for diploid species) is called **polyploid**. For instance, fertilization of an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely rare, with only a few examples among the flatworms, crustaceans, amphibians, fish, and lizards. Triploid animals are sterile because meiosis cannot proceed normally with an odd number of chromosome sets. In contrast, polyploidy is very common in the plant kingdom, and polyploid plants tend to be larger and more robust than euploids of their species.

Chromosome Structural Rearrangements

Cytologists have characterized numerous structural rearrangements in chromosomes, including partial duplications, deletions, inversions, and translocations. Duplications and deletions often produce offspring that survive but exhibit physical and mental abnormalities. Cri-du-chat (from the French for “cry of the cat”) is a syndrome associated with nervous system abnormalities and identifiable physical features that results from a deletion of most of the small arm of chromosome 5 ([link]). Infants with this genotype emit a characteristic high-pitched cry upon which the disorder’s name is based.



This individual with cri-du-chat syndrome is shown at various ages: (A) age two, (B) age four, (C) age

nine, and (D) age 12. (credit: Paola Cerruti Mainardi)

Chromosome inversions and translocations can be identified by observing cells during meiosis because homologous chromosomes with a rearrangement in one of the pair must contort to maintain appropriate gene alignment and pair effectively during prophase I.

A **chromosome inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome ([link]). Unless they disrupt a gene sequence, inversions only change the orientation of genes and are likely to have more mild effects than aneuploid errors.

Note:
Evolution in Action
The Chromosome 18 Inversion
Not all structural rearrangements of chromosomes produce nonviable, impaired, or infertile individuals. In rare instances, such a change can result in the evolution of a new species. In fact, an inversion in chromosome 18 appears to have contributed to the evolution of humans. This inversion is not present in our closest genetic relatives, the chimpanzees.

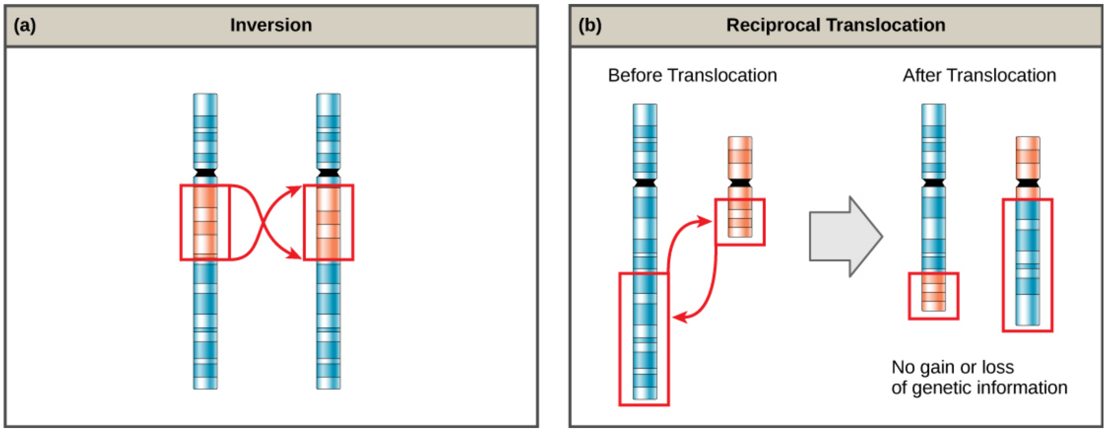
The chromosome 18 inversion is believed to have occurred in early humans following their divergence from a common ancestor with chimpanzees approximately five million years ago. Researchers have suggested that a long stretch of DNA was duplicated on chromosome 18 of an ancestor to humans, but that during the duplication it was inverted (inserted into the chromosome in reverse orientation.

A comparison of human and chimpanzee genes in the region of this inversion indicates that two genes—*ROCK1* and *USP14*—are farther apart on human chromosome 18 than they are on the corresponding chimpanzee chromosome. This suggests that one of the inversion breakpoints occurred between these two genes. Interestingly, humans and chimpanzees express *USP14* at distinct levels in specific cell types, including cortical cells and fibroblasts. Perhaps the chromosome 18 inversion in an ancestral human

repositioned specific genes and reset their expression levels in a useful way. Because both *ROCK1* and *USP14* code for enzymes, a change in their expression could alter cellular function. It is not known how this inversion contributed to hominid evolution, but it appears to be a significant factor in the divergence of humans from other primates.[footnote]
V Goidts, et al., “Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates,” *Human Genetics*, 115 (2004):116–22.

A translocation occurs when a segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome. Translocations can be benign or have devastating effects, depending on how the positions of genes are altered with respect to regulatory sequences. Notably, specific translocations have been associated with several cancers and with schizophrenia. Reciprocal translocations result from the exchange of chromosome segments between two nonhomologous chromosomes such that there is no gain or loss of genetic information ([link]).

An (a) inversion occurs when a chromosome segment breaks from the chromosome, reverses its orientation, and then reattaches in the original position. A (b) reciprocal translocation occurs between two nonhomologous chromosomes and does not cause any genetic information to be lost or duplicated. (credit: modification of work by National Human Genome Research Institute (USA)



Section Summary

The number, size, shape, and banding pattern of chromosomes make them easily identifiable in a karyogram and allow for the assessment of many chromosomal abnormalities. Disorders in chromosome number, or aneuploidies, are typically lethal to the embryo, although a few trisomic genotypes are viable. Because of X inactivation, aberrations in sex chromosomes typically have milder effects on an individual. Aneuploidies also include instances in which segments of a chromosome are duplicated or deleted. Chromosome structures also may be rearranged, for example by inversion or translocation. Both of these aberrations can result in negative effects on development, or death. Because they force chromosomes to assume contorted pairings during meiosis I, inversions and translocations are often associated with reduced fertility because of the likelihood of nondisjunction.

Multiple Choice

Exercise:

Problem: The genotype XXY corresponds to:

a. Klinefelter syndrome
b. Turner syndrome
c. Triplo-X
d. Jacob syndrome

Solution:

A

Exercise:

Problem:

Abnormalities in the number of X chromosomes tend to be milder than the same abnormalities in autosomes because of _________.

a. deletions
b. nonhomologous recombination
c. synapsis
d. X inactivation

Solution:

D

Exercise:

Problem:

Aneuploidies are deleterious for the individual because of what phenomenon?

a. nondisjunction
b. gene dosage
c. meiotic errors
d. X inactivation

Solution:

B

Free Response

Exercise:

Problem:

Individuals with trisomy 21 are more likely to survive to adulthood than individuals with trisomy 18. Based on what you know about aneuploidies from this module, what can you hypothesize about chromosomes 21 and 18?

Solution:

The problems caused by trisomies arise because the genes on the chromosome that is present in three copies produce more product than genes on chromosomes with only two copies. The cell does not have a way to adjust the amount of product, and the lack of balance causes problems in development and the maintenance of the individual. Each chromosome is different, and the differences in survivability could have to do with the numbers of genes on the two chromosomes. Chromosome 21 may be a smaller chromosome, so there are fewer unbalanced gene products. It is also possible that chromosome 21 carries genes whose products are less sensitive to differences in dosage than chromosome 18. The genes may be less involved in critical pathways, or the differences in dosage may make less of a difference to those pathways.

Glossary

aneuploid
: an individual with an error in chromosome number; includes deletions and duplications of chromosome segments

autosome
: any of the non-sex chromosomes

chromosome inversion

the detachment, 180° rotation, and reinsertion of a chromosome arm

euploid
: an individual with the appropriate number of chromosomes for their species

karyogram
: the photographic image of a karyotype

karyotype
: the number and appearance of an individuals chromosomes, including the size, banding patterns, and centromere position

monosomy
: an otherwise diploid genotype in which one chromosome is missing

nondisjunction
: the failure of synapsed homologs to completely separate and migrate to separate poles during the first cell division of meiosis

polyploid
: an individual with an incorrect number of chromosome sets

translocation
: the process by which one segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome

trisomy
: an otherwise diploid genotype in which one entire chromosome is duplicated

X inactivation
: the condensation of X chromosomes into Barr bodies during embryonic development in females to compensate for the double genetic dose

Introduction
class="introduction"

Experimenting with thousands of garden peas, Mendel uncovered the fundamentals of genetics. (credit: modification of work by Jerry Kirkhart)

Genetics is the study of heredity. Johann Gregor Mendel set the framework for genetics long before chromosomes or genes had been identified, at a time when meiosis was not well understood. Mendel selected a simple biological system and conducted methodical, quantitative analyses using large sample sizes. Because of Mendel’s work, the fundamental principles of heredity were revealed. We now know that genes, carried on chromosomes, are the basic functional units of heredity with the ability to be replicated, expressed, or mutated. Today, the postulates put forth by Mendel form the basis of classical, or Mendelian, genetics. Not all genes are transmitted from parents to offspring according to Mendelian genetics, but Mendel’s experiments serve as an excellent starting point for thinking about inheritance.

Mendel's Experiments

By the end of this section, you will be able to:

- Explain the scientific reasons for the success of Mendel's experimental work
- Describe the expected outcomes of monohybrid crosses involving dominant and recessive alleles

Johann Gregor Mendel set the framework for the study of genetics.

Johann Gregor Mendel (1822–1884) ([link]) was a lifelong learner, teacher, scientist, and man of faith. As a young adult, he joined the Augustinian Abbey of St. Thomas in Brno in what is now the Czech Republic. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels. In 1856, he began a decade-long research pursuit involving inheritance patterns in honeybees and plants, ultimately settling on pea plants as his primary **model system** (a system with convenient characteristics that is used to study a specific biological phenomenon to gain understanding to be applied to other systems). In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local natural history society. He

demonstrated that traits are transmitted faithfully from parents to offspring in specific patterns. In 1866, he published his work, *Experiments in Plant Hybridization,*[footnote] in the proceedings of the Natural History Society of Brünn.

Johann Gregor Mendel, “Versuche über Pflanzenhybriden.” *Verhandlungen des naturforschenden Vereines in Brünn*, Bd. IV für das Jahr, 1865 Abhandlungen (1866):3–47. [for English translation, see http://www.mendelweb.org/Mendel.plain.html]

Mendel’s work went virtually unnoticed by the scientific community, which incorrectly believed that the process of inheritance involved a blending of parental traits that produced an intermediate physical appearance in offspring. This hypothetical process appeared to be correct because of what we know now as continuous variation. **Continuous variation** is the range of small differences we see among individuals in a characteristic like human height. It does appear that offspring are a “blend” of their parents’ traits when we look at characteristics that exhibit continuous variation. Mendel worked instead with traits that show **discontinuous variation**. Discontinuous variation is the variation seen among individuals when each individual shows one of two—or a very few—easily distinguishable traits, such as violet or white flowers. Mendel’s choice of these kinds of traits allowed him to see experimentally that the traits were not blended in the offspring as would have been expected at the time, but that they were inherited as distinct traits. In 1868, Mendel became abbot of the monastery and exchanged his scientific pursuits for his pastoral duties. He was not recognized for his extraordinary scientific contributions during his lifetime; in fact, it was not until 1900 that his work was rediscovered, reproduced, and revitalized by scientists on the brink of discovering the chromosomal basis of heredity.

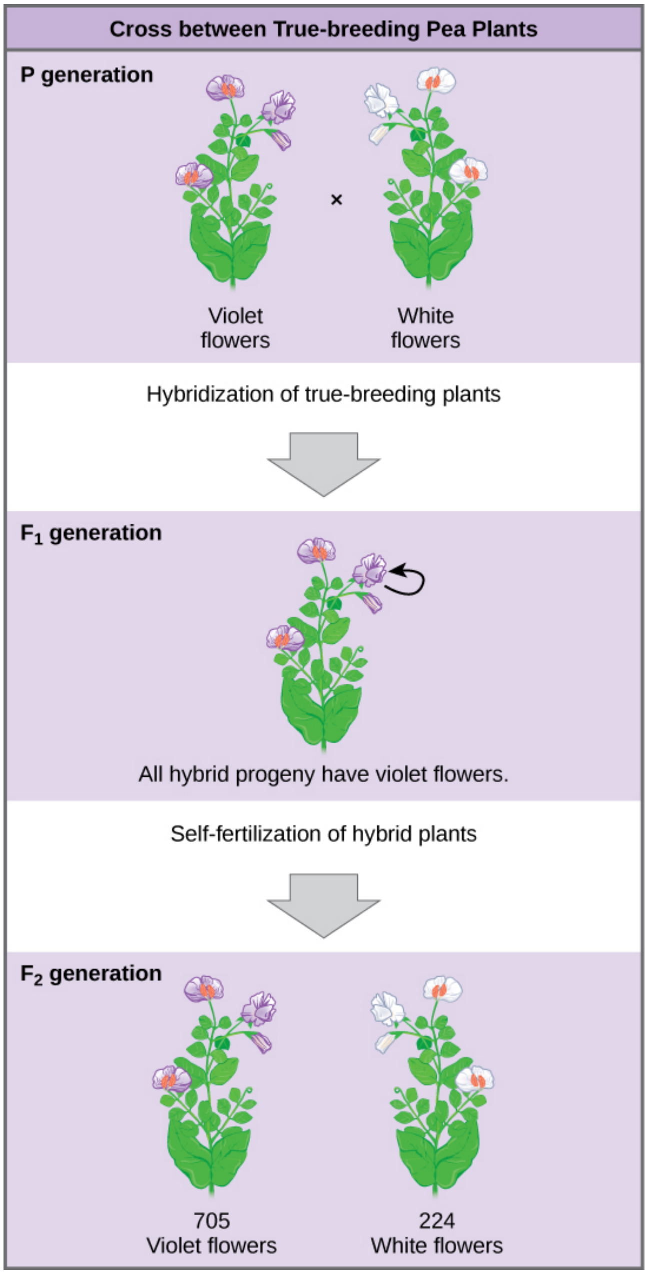
Mendel’s Crosses

Mendel’s seminal work was accomplished using the garden pea, *Pisum sativum*, to study inheritance. This species naturally self-fertilizes, meaning that pollen encounters ova within the same flower. The flower petals remain sealed tightly until pollination is completed to prevent the pollination of other plants. The result is highly inbred, or “true-breeding,” pea plants.

These are plants that always produce offspring that look like the parent. By experimenting with true-breeding pea plants, Mendel avoided the appearance of unexpected traits in offspring that might occur if the plants were not true breeding. The garden pea also grows to maturity within one season, meaning that several generations could be evaluated over a relatively short time. Finally, large quantities of garden peas could be cultivated simultaneously, allowing Mendel to conclude that his results did not come about simply by chance.

Mendel performed **hybridizations**, which involve mating two true-breeding individuals that have different traits. In the pea, which is naturally self-pollinating, this is done by manually transferring pollen from the anther of a mature pea plant of one variety to the stigma of a separate mature pea plant of the second variety.

Plants used in first-generation crosses were called **P**, or parental generation, plants ([link]). Mendel collected the seeds produced by the P plants that resulted from each cross and grew them the following season. These offspring were called the $\mathbf{F_1}$, or the first filial (filial = daughter or son), generation. Once Mendel examined the characteristics in the F_1 generation of plants, he allowed them to self-fertilize naturally. He then collected and grew the seeds from the F_1 plants to produce the $\mathbf{F_2}$, or second filial, generation. Mendel's experiments extended beyond the F_2 generation to the F_3 generation, F_4 generation, and so on, but it was the ratio of characteristics in the P, F_1, and F_2 generations that were the most intriguing and became the basis of Mendel's postulates.



Mendel's process for performing crosses included examining flower color.

Garden Pea Characteristics Revealed the Basics of Heredity

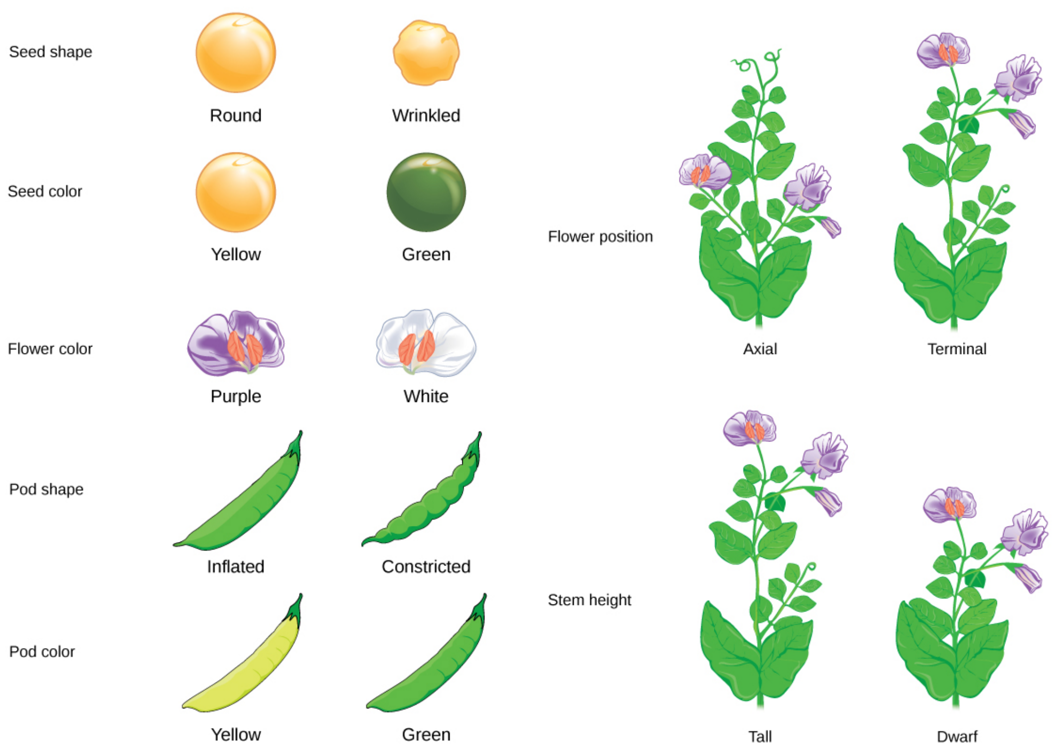
In his 1865 publication, Mendel reported the results of his crosses involving seven different characteristics, each with two contrasting traits. A **trait** is defined as a variation in the physical appearance of a heritable characteristic. The characteristics included plant height, seed texture, seed color, flower color, pea-pod size, pea-pod color, and flower position. For the characteristic of flower color, for example, the two contrasting traits were white versus violet. To fully examine each characteristic, Mendel generated large numbers of F_1 and F_2 plants and reported results from thousands of F_2 plants.

What results did Mendel find in his crosses for flower color? First, Mendel confirmed that he was using plants that bred true for white or violet flower color. Irrespective of the number of generations that Mendel examined, all self-crossed offspring of parents with white flowers had white flowers, and all self-crossed offspring of parents with violet flowers had violet flowers. In addition, Mendel confirmed that, other than flower color, the pea plants were physically identical. This was an important check to make sure that the two varieties of pea plants only differed with respect to one trait, flower color.

Once these validations were complete, Mendel applied the pollen from a plant with violet flowers to the stigma of a plant with white flowers. After gathering and sowing the seeds that resulted from this cross, Mendel found that 100 percent of the F_1 hybrid generation had violet flowers. Conventional wisdom at that time would have predicted the hybrid flowers to be pale violet or for hybrid plants to have equal numbers of white and violet flowers. In other words, the contrasting parental traits were expected to blend in the offspring. Instead, Mendel's results demonstrated that the white flower trait had completely disappeared in the F_1 generation.

Importantly, Mendel did not stop his experimentation there. He allowed the F_1 plants to self-fertilize and found that 705 plants in the F_2 generation had violet flowers and 224 had white flowers. This was a ratio of 3.15 violet flowers to one white flower, or approximately 3:1. When Mendel transferred pollen from a plant with violet flowers to the stigma of a plant with white flowers and vice versa, he obtained approximately the same ratio irrespective of which parent—male or female—contributed which trait.

This is called a **reciprocal cross**—a paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross. For the other six characteristics that Mendel examined, the F_1 and F_2 generations behaved in the same way that they behaved for flower color. One of the two traits would disappear completely from the F_1 generation, only to reappear in the F_2 generation at a ratio of roughly 3:1 ([link]).



Mendel identified seven pea plant characteristics.

Upon compiling his results for many thousands of plants, Mendel concluded that the characteristics could be divided into expressed and latent traits. He called these dominant and recessive traits, respectively. **Dominant** traits are those that are inherited unchanged in a hybridization. **Recessive** traits become latent, or disappear in the offspring of a hybridization. The recessive trait does, however, reappear in the progeny of

the hybrid offspring. An example of a dominant trait is the violet-colored flower trait. For this same characteristic (flower color), white-colored flowers are a recessive trait. The fact that the recessive trait reappeared in the F_2 generation meant that the traits remained separate (and were not blended) in the plants of the F_1 generation. Mendel proposed that this was because the plants possessed two copies of the trait for the flower-color characteristic, and that each parent transmitted one of their two copies to their offspring, where they came together. Moreover, the physical observation of a dominant trait could mean that the genetic composition of the organism included two dominant versions of the characteristic, or that it included one dominant and one recessive version. Conversely, the observation of a recessive trait meant that the organism lacked any dominant versions of this characteristic.

Note:
Concept in Action

For an excellent review of Mendel's experiments and to perform your own crosses and identify patterns of inheritance, visit the Mendel's Peas web lab.

Section Summary

Working with garden pea plants, Mendel found that crosses between parents that differed for one trait produced F_1 offspring that all expressed one parent's traits. The traits that were visible in the F_1 generation are referred to as dominant, and traits that disappear in the F_1 generation are described as recessive. When the F_1 plants in Mendel's experiment were self-crossed,

the F_2 offspring exhibited the dominant trait or the recessive trait in a 3:1 ratio, confirming that the recessive trait had been transmitted faithfully from the original P parent. Reciprocal crosses generated identical F_1 and F_2 offspring ratios. By examining sample sizes, Mendel showed that traits were inherited as independent events.

Multiple Choice

Exercise:

Problem:

Imagine that you are performing a cross involving seed color in garden pea plants. What traits would you expect to observe in the F_1 offspring if you cross true-breeding parents with green seeds and yellow seeds? Yellow seed color is dominant over green.

a. only yellow-green seeds
b. only yellow seeds
c. 1:1 yellow seeds:green seeds
d. 1:3 green seeds:yellow seeds

Solution:

B

Exercise:

Problem:

Imagine that you are performing a cross involving seed texture in garden pea plants. You cross true-breeding round and wrinkled parents to obtain F_1 offspring. Which of the following experimental results in terms of numbers of plants are closest to what you expect in the F_2 progeny?

a. 810 round seeds
b. 810 wrinkled seeds

c. 405:395 round seeds:wrinkled seeds
d. 610:190 round seeds:wrinkled seeds

Solution:

D

Free Response

Exercise:

Problem:

Describe one of the reasons that made the garden pea an excellent choice of model system for studying inheritance.

Solution:

The garden pea has flowers that close tightly during self-pollination. This helps to prevent accidental or unintentional fertilizations that could have diminished the accuracy of Mendel’s data.

Glossary

continuous variation
: a variation in a characteristic in which individuals show a range of traits with small differences between them

discontinuous variation
: a variation in a characteristic in which individuals show two, or a few, traits with large differences between them

dominant
: describes a trait that masks the expression of another trait when both versions of the gene are present in an individual

F_1
: the first filial generation in a cross; the offspring of the parental generation

F_2
: the second filial generation produced when F_1 individuals are self-crossed or fertilized with each other

hybridization
: the process of mating two individuals that differ, with the goal of achieving a certain characteristic in their offspring

model system
: a species or biological system used to study a specific biological phenomenon to gain understanding that will be applied to other species

P
: the parental generation in a cross

recessive
: describes a trait whose expression is masked by another trait when the alleles for both traits are present in an individual

reciprocal cross
: a paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross

trait
: a variation in an inherited characteristic

Laws of Inheritance
By the end of this section, you will be able to:

- Explain the relationship between genotypes and phenotypes in dominant and recessive gene systems
- Use a Punnett square to calculate the expected proportions of genotypes and phenotypes in a monohybrid cross
- Explain Mendel’s law of segregation and independent assortment in terms of genetics and the events of meiosis
- Explain the purpose and methods of a test cross

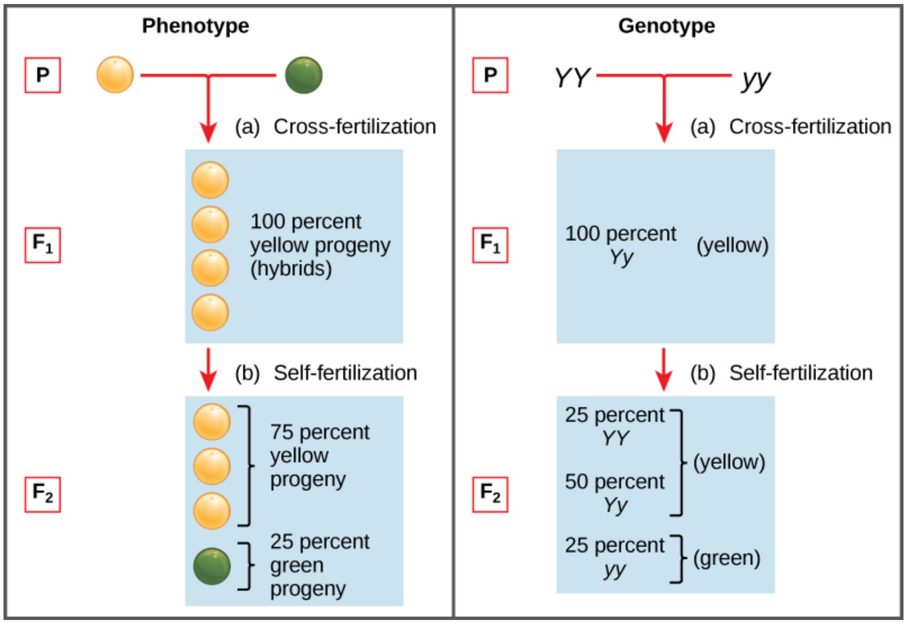
The seven characteristics that Mendel evaluated in his pea plants were each expressed as one of two versions, or traits. Mendel deduced from his results that each individual had two discrete copies of the characteristic that are passed individually to offspring. We now call those two copies genes, which are carried on chromosomes. The reason we have two copies of each gene is that we inherit one from each parent. In fact, it is the chromosomes we inherit and the two copies of each gene are located on paired chromosomes. Recall that in meiosis these chromosomes are separated out into haploid gametes. This separation, or segregation, of the homologous chromosomes means also that only one of the copies of the gene gets moved into a gamete. The offspring are formed when that gamete unites with one from another parent and the two copies of each gene (and chromosome) are restored.

For cases in which a single gene controls a single characteristic, a diploid organism has two genetic copies that may or may not encode the same version of that characteristic. For example, one individual may carry a gene that determines white flower color and a gene that determines violet flower color. Gene variants that arise by mutation and exist at the same relative locations on homologous chromosomes are called **alleles**. Mendel examined the inheritance of genes with just two allele forms, but it is common to encounter more than two alleles for any given gene in a natural population.

Phenotypes and Genotypes

Two alleles for a given gene in a diploid organism are expressed and interact to produce physical characteristics. The observable traits expressed by an organism are referred to as its **phenotype**. An organism's underlying genetic makeup, consisting of both the physically visible and the non-expressed alleles, is called its **genotype**. Mendel's hybridization experiments demonstrate the difference between phenotype and genotype. For example, the phenotypes that Mendel observed in his crosses between pea plants with differing traits are connected to the diploid genotypes of the plants in the P, F_1, and F_2 generations. We will use a second trait that Mendel investigated, seed color, as an example. Seed color is governed by a single gene with two alleles. The yellow-seed allele is dominant and the green-seed allele is recessive. When true-breeding plants were cross-fertilized, in which one parent had yellow seeds and one had green seeds, all of the F_1 hybrid offspring had yellow seeds. That is, the hybrid offspring were phenotypically identical to the true-breeding parent with yellow seeds. However, we know that the allele donated by the parent with green seeds was not simply lost because it reappeared in some of the F_2 offspring ([link]). Therefore, the F_1 plants must have been genotypically different from the parent with yellow seeds.

The P plants that Mendel used in his experiments were each homozygous for the trait he was studying. Diploid organisms that are **homozygous** for a gene have two identical alleles, one on each of their homologous chromosomes. The genotype is often written as *YY* or *yy*, for which each letter represents one of the two alleles in the genotype. The dominant allele is capitalized and the recessive allele is lower case. The letter used for the gene (seed color in this case) is usually related to the dominant trait (yellow allele, in this case, or "*Y*"). Mendel's parental pea plants always bred true because both produced gametes carried the same allele. When P plants with contrasting traits were cross-fertilized, all of the offspring were **heterozygous** for the contrasting trait, meaning their genotype had different alleles for the gene being examined. For example, the F_1 yellow plants that received a *Y* allele from their yellow parent and a *y* allele from their green parent had the genotype *Yy*.



Phenotypes are physical expressions of traits that are transmitted by alleles. Capital letters represent dominant alleles and lowercase letters represent recessive alleles. The phenotypic ratios are the ratios of visible characteristics. The genotypic ratios are the ratios of gene combinations in the offspring, and these are not always distinguishable in the phenotypes.

Law of Dominance

Our discussion of homozygous and heterozygous organisms brings us to why the F_1 heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant, and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosomes. For a gene that is expressed in a dominant and recessive

pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype), and the recessive allele will only be observed in homozygous recessive individuals ([link]).

Correspondence between Genotype and Phenotype for a Dominant-Recessive Characteristic.			
	Homozygous	**Heterozygous**	**Homozygous**
Genotype	*YY*	*Yy*	*yy*
Phenotype	yellow	yellow	green

Mendel’s **law of dominance** states that in a heterozygote, one trait will conceal the presence of another trait for the same characteristic. For example, when crossing true-breeding violet-flowered plants with true-breeding white-flowered plants, all of the offspring were violet-flowered, even though they all had one allele for violet and one allele for white. Rather than both alleles contributing to a phenotype, the dominant allele will be expressed exclusively. The recessive allele will remain latent, but will be transmitted to offspring in the same manner as that by which the dominant allele is transmitted. The recessive trait will only be expressed by offspring that have two copies of this allele ([link]), and these offspring will breed true when self-crossed.

The allele for albinism, expressed here in humans, is recessive. Both of this child’s parents carried the recessive allele.

Monohybrid Cross and the Punnett Square

When fertilization occurs between two true-breeding parents that differ by only the characteristic being studied, the process is called a **monohybrid** cross, and the resulting offspring are called monohybrids. Mendel performed seven types of monohybrid crosses, each involving contrasting traits for different characteristics. Out of these crosses, all of the F_1 offspring had the phenotype of one parent, and the F_2 offspring had a 3:1 phenotypic ratio. On the basis of these results, Mendel postulated that each parent in the monohybrid cross contributed one of two paired unit factors to each offspring, and every possible combination of unit factors was equally likely.

The results of Mendel’s research can be explained in terms of probabilities, which are mathematical measures of likelihood. The probability of an event is calculated by the number of times the event occurs divided by the total number of opportunities for the event to occur. A probability of one (100 percent) for some event indicates that it is guaranteed to occur, whereas a probability of zero (0 percent) indicates that it is guaranteed to not occur, and a probability of 0.5 (50 percent) means it has an equal chance of occurring or not occurring.

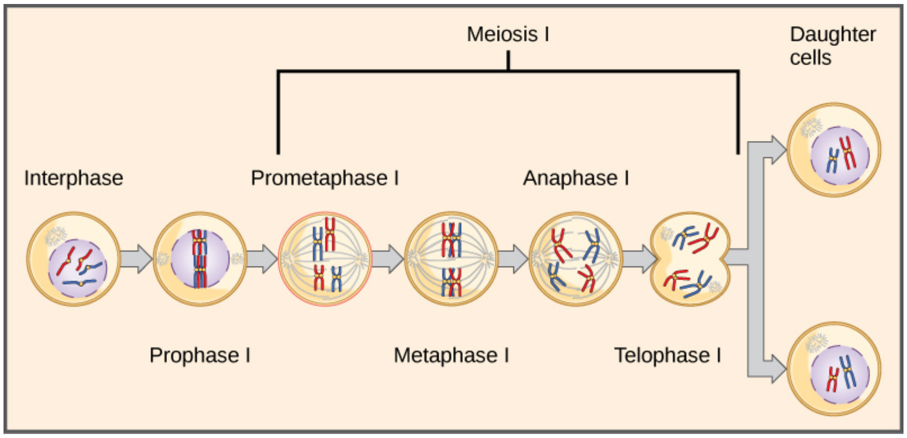
To demonstrate this with a monohybrid cross, consider the case of true-breeding pea plants with yellow versus green seeds. The dominant seed color is yellow; therefore, the parental genotypes were *YY* for the plants with yellow seeds and *yy* for the plants with green seeds. A **Punnett square**, devised by the British geneticist Reginald Punnett, is useful for determining probabilities because it is drawn to predict all possible outcomes of all possible random fertilization events and their expected frequencies. [link] shows a Punnett square for a cross between a plant with yellow peas and one with green peas. To prepare a Punnett square, all possible combinations of the parental alleles (the genotypes of the gametes) are listed along the top (for one parent) and side (for the other parent) of a grid. The combinations of egg and sperm gametes are then made in the boxes in the table on the basis of which alleles are combining. Each box then represents the diploid genotype of a zygote, or fertilized egg. Because each possibility is equally likely, genotypic ratios can be determined from a Punnett square. If the pattern of inheritance (dominant and recessive) is known, the phenotypic ratios can be inferred as well. For a monohybrid cross of two true-breeding parents, each parent contributes one type of allele. In this case, only one genotype is possible in the F_1 offspring. All offspring are *Yy* and have yellow seeds.

When the F_1 offspring are crossed with each other, each has an equal probability of contributing either a *Y* or a *y* to the F_2 offspring. The result is a 1 in 4 (25 percent) probability of both parents contributing a *Y*, resulting in an offspring with a yellow phenotype; a 25 percent probability of parent A contributing a *Y* and parent B a *y*, resulting in offspring with a yellow phenotype; a 25 percent probability of parent A contributing a *y* and parent B a *Y*, also resulting in a yellow phenotype; and a (25 percent) probability

of both parents contributing a *y*, resulting in a green phenotype. When counting all four possible outcomes, there is a 3 in 4 probability of offspring having the yellow phenotype and a 1 in 4 probability of offspring having the green phenotype. This explains why the results of Mendel's F_2 generation occurred in a 3:1 phenotypic ratio. Using large numbers of crosses, Mendel was able to calculate probabilities, found that they fit the model of inheritance, and use these to predict the outcomes of other crosses.

Law of Segregation

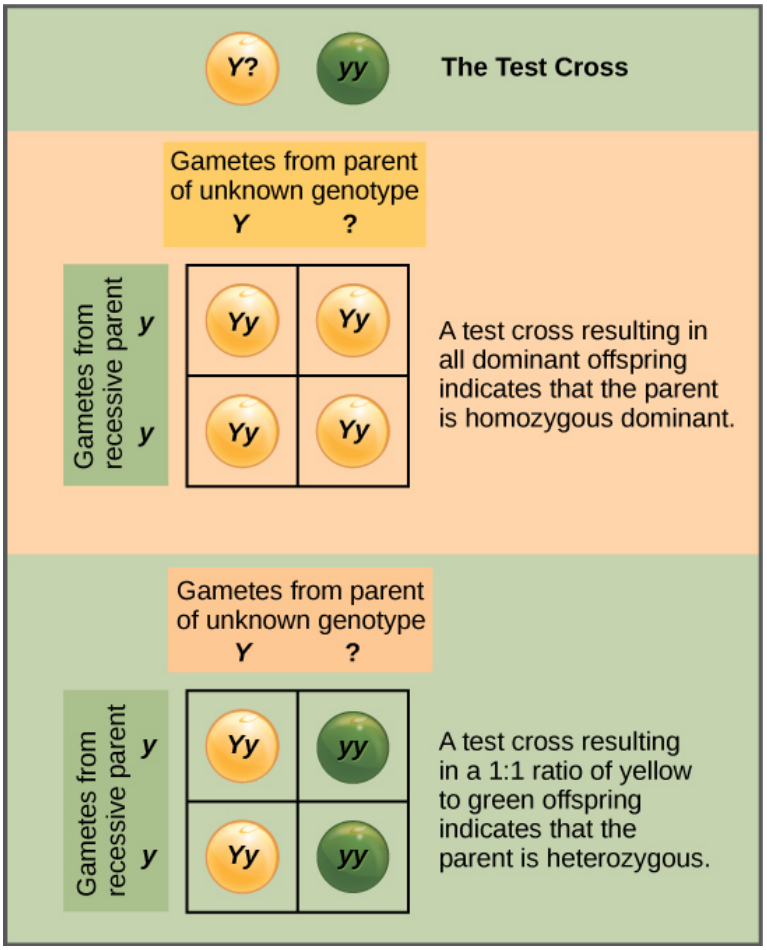
Observing that true-breeding pea plants with contrasting traits gave rise to F_1 generations that all expressed the dominant trait and F_2 generations that expressed the dominant and recessive traits in a 3:1 ratio, Mendel proposed the **law of segregation**. This law states that paired unit factors (genes) must segregate equally into gametes such that offspring have an equal likelihood of inheriting either factor. For the F_2 generation of a monohybrid cross, the following three possible combinations of genotypes result: homozygous dominant, heterozygous, or homozygous recessive. Because heterozygotes could arise from two different pathways (receiving one dominant and one recessive allele from either parent), and because heterozygotes and homozygous dominant individuals are phenotypically identical, the law supports Mendel's observed 3:1 phenotypic ratio. The equal segregation of alleles is the reason we can apply the Punnett square to accurately predict the offspring of parents with known genotypes. The physical basis of Mendel's law of segregation is the first division of meiosis in which the homologous chromosomes with their different versions of each gene are segregated into daughter nuclei. This process was not understood by the scientific community during Mendel's lifetime ([link]).



The first division in meiosis is shown.

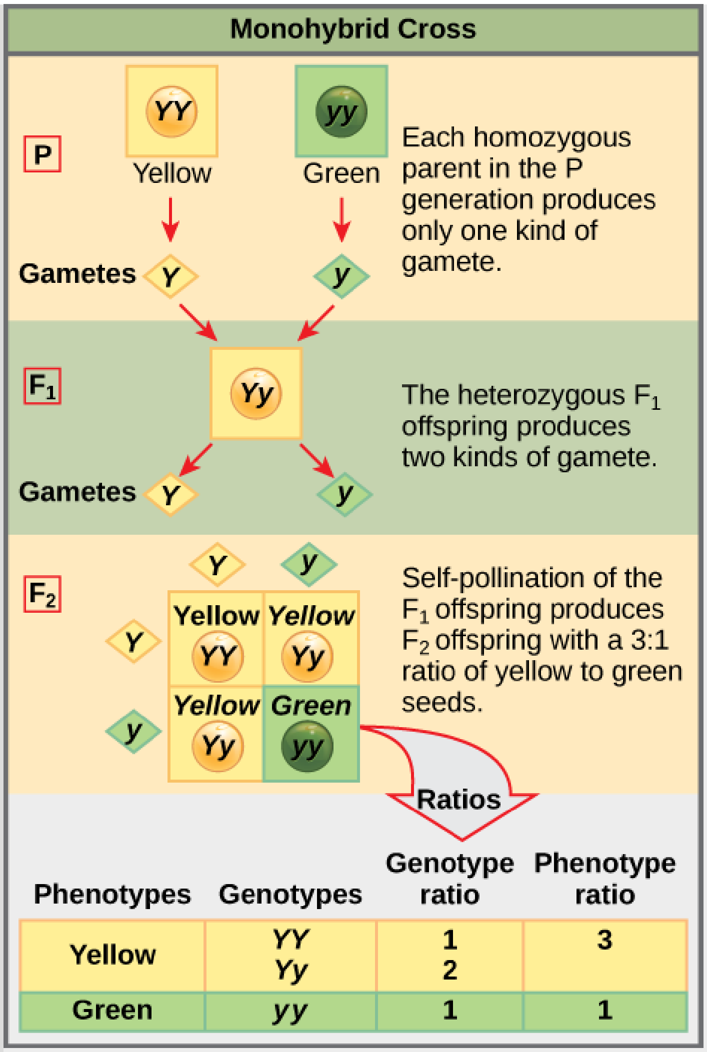
Test Cross

Beyond predicting the offspring of a cross between known homozygous or heterozygous parents, Mendel also developed a way to determine whether an organism that expressed a dominant trait was a heterozygote or a homozygote. Called the **test cross**, this technique is still used by plant and animal breeders. In a test cross, the dominant-expressing organism is crossed with an organism that is homozygous recessive for the same characteristic. If the dominant-expressing organism is a homozygote, then all F_1 offspring will be heterozygotes expressing the dominant trait ([link]). Alternatively, if the dominant-expressing organism is a heterozygote, the F_1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes ([link]). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.



A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

Note:
Art Connection



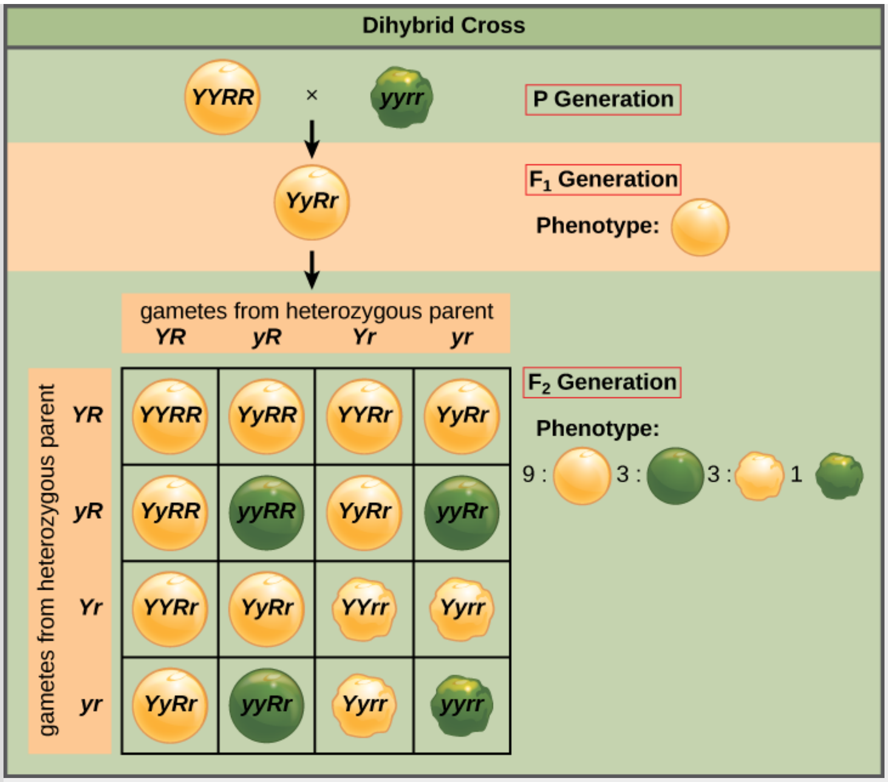
This Punnett square shows the cross between plants with yellow seeds and green seeds. The cross between the true-breeding P plants produces F_1 heterozygotes that can be self-fertilized. The self-cross of the F_1 generation can be analyzed with a Punnett square to predict the genotypes of the F_2 generation. Given an inheritance pattern of dominant–recessive, the genotypic and phenotypic ratios can then be determined.

In pea plants, round peas (*R*) are dominant to wrinkled peas (*r*). You do a test cross between a pea plant with wrinkled peas (genotype *rr*) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas. From this data, can you tell if the parent plant is homozygous dominant or heterozygous?

Law of Independent Assortment

Mendel's **law of independent assortment** states that genes do not influence each other with regard to the sorting of alleles into gametes, and every possible combination of alleles for every gene is equally likely to occur. Independent assortment of genes can be illustrated by the **dihybrid** cross, a cross between two true-breeding parents that express different traits for two characteristics. Consider the characteristics of seed color and seed texture for two pea plants, one that has wrinkled, green seeds (*rryy*) and another that has round, yellow seeds (*RRYY*). Because each parent is homozygous, the law of segregation indicates that the gametes for the wrinkled–green plant all are *ry*, and the gametes for the round–yellow plant are all *RY*. Therefore, the F_1 generation of offspring all are *RrYy* ([link]).

Note:
Art Connection



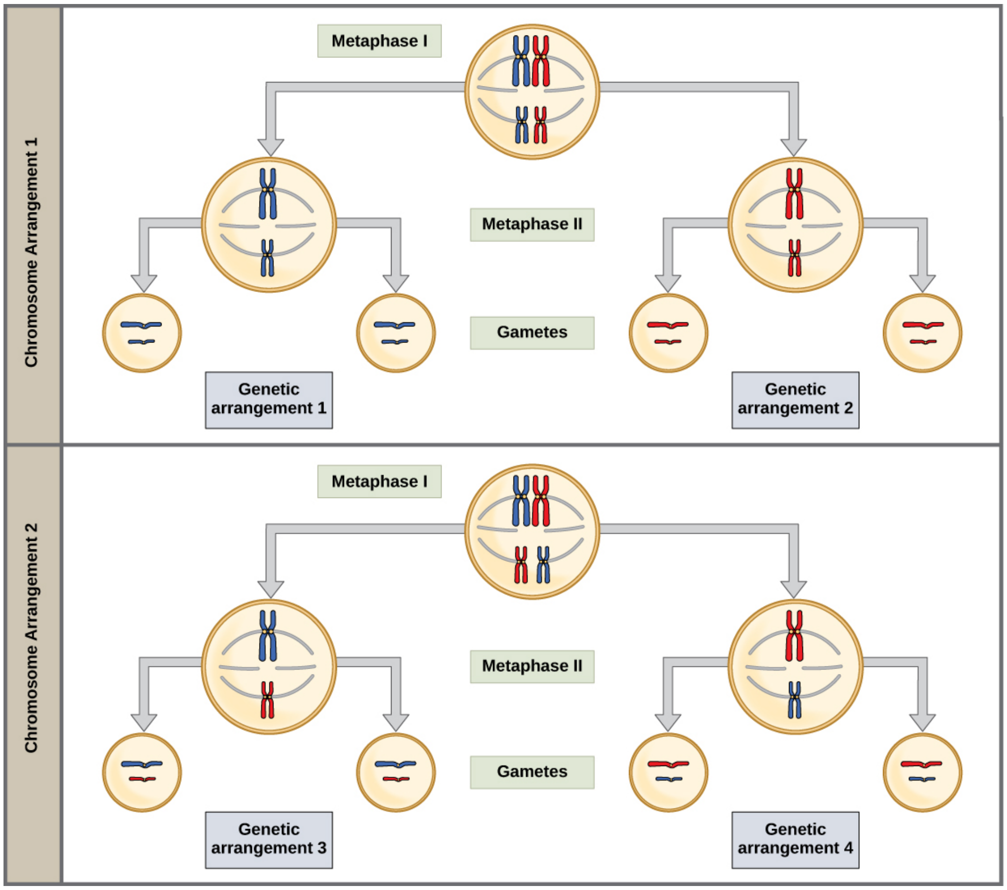
A dihybrid cross in pea plants involves the genes for seed color and texture. The P cross produces F_1 offspring that are all heterozygous for both characteristics. The resulting 9:3:3:1 F_2 phenotypic ratio is obtained using a Punnett square.

In pea plants, purple flowers (*P*) are dominant to white (*p*), and yellow peas (*Y*) are dominant to green (*y*). What are the possible genotypes and phenotypes for a cross between *PpYY* and *ppYy* pea plants? How many squares would you need to complete a Punnett square analysis of this cross?

The gametes produced by the F_1 individuals must have one allele from each of the two genes. For example, a gamete could get an *R* allele for the seed shape gene and either a *Y* or a *y* allele for the seed color gene. It cannot get

both an *R* and an *r* allele; each gamete can have only one allele per gene. The law of independent assortment states that a gamete into which an *r* allele is sorted would be equally likely to contain either a *Y* or a *y* allele. Thus, there are four equally likely gametes that can be formed when the *RrYy* heterozygote is self-crossed, as follows: *RY*, *rY*, *Ry*, and *ry*. Arranging these gametes along the top and left of a 4 × 4 Punnett square ([link]) gives us 16 equally likely genotypic combinations. From these genotypes, we find a phenotypic ratio of 9 round–yellow:3 round–green:3 wrinkled–yellow:1 wrinkled–green ([link]). These are the offspring ratios we would expect, assuming we performed the crosses with a large enough sample size.

The physical basis for the law of independent assortment also lies in meiosis I, in which the different homologous pairs line up in random orientations. Each gamete can contain any combination of paternal and maternal chromosomes (and therefore the genes on them) because the orientation of tetrads on the metaphase plane is random ([link]).



The random segregation into daughter nuclei that happens during the first division in meiosis can lead to a variety of possible genetic arrangements.

Section Summary

When true-breeding, or homozygous, individuals that differ for a certain trait are crossed, all of the offspring will be heterozygous for that trait. If the traits are inherited as dominant and recessive, the F_1 offspring will all exhibit the same phenotype as the parent homozygous for the dominant trait. If these heterozygous offspring are self-crossed, the resulting F_2 offspring will be equally likely to inherit gametes carrying the dominant or recessive trait, giving rise to offspring of which one quarter are homozygous dominant, half are heterozygous, and one quarter are

homozygous recessive. Because homozygous dominant and heterozygous individuals are phenotypically identical, the observed traits in the F_2 offspring will exhibit a ratio of three dominant to one recessive.

Mendel postulated that genes (characteristics) are inherited as pairs of alleles (traits) that behave in a dominant and recessive pattern. Alleles segregate into gametes such that each gamete is equally likely to receive either one of the two alleles present in a diploid individual. In addition, genes are assorted into gametes independently of one another. That is, in general, alleles are not more likely to segregate into a gamete with a particular allele of another gene.

Art Connections

Exercise:

Problem:

[link] In pea plants, round peas (*R*) are dominant to wrinkled peas (*r*). You do a test cross between a pea plant with wrinkled peas (genotype *rr*) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas. From this data, can you tell if the parent plant is homozygous dominant or heterozygous?

Solution:

[link] You cannot be sure if the plant is homozygous or heterozygous as the data set is too small: by random chance, all three plants might have acquired only the dominant gene even if the recessive one is present.

Exercise:

Problem:

[link] In pea plants, purple flowers (*P*) are dominant to white (*p*), and yellow peas (*Y*) are dominant to green (*y*). What are the possible genotypes and phenotypes for a cross between *PpYY* and *ppYy* pea plants? How many squares would you need to complete a Punnett square analysis of this cross?

Solution:

[link] The possible genotypes are *PpYY*, *PpYy*, *ppYY*, and *ppYy*. The former two genotypes would result in plants with purple flowers and yellow peas, while the latter two genotypes would result in plants with white flowers with yellow peas, for a 1:1 ratio of each phenotype. You only need a 2 × 2 Punnett square (four squares total) to do this analysis because two of the alleles are homozygous.

Multiple Choice

Exercise:

Problem:

The observable traits expressed by an organism are described as its _________.

a. phenotype
b. genotype
c. alleles
d. zygote

Solution:

A

Exercise:

Problem:

A recessive trait will be observed in individuals that are _________ for that trait.

a. heterozygous
b. homozygous or heterozygous
c. homozygous
d. diploid

Solution:

C

Exercise:

Problem:

What are the types of gametes that can be produced by an individual with the genotype *AaBb*?

a. *Aa*, *Bb*
b. *AA*, *aa*, *BB*, *bb*
c. *AB*, *Ab*, *aB*, *ab*
d. *AB*, *ab*

Solution:

C

Exercise:

Problem: What is the reason for doing a test cross?

a. to identify heterozygous individuals with the dominant phenotype
b. to determine which allele is dominant and which is recessive
c. to identify homozygous recessive individuals in the F_2

d. to determine if two genes assort independently

Solution:

A

Free Response

Exercise:

Problem:

Use a Punnett square to predict the offspring in a cross between a dwarf pea plant (homozygous recessive) and a tall pea plant (heterozygous). What is the phenotypic ratio of the offspring?

Solution:

The Punnett square would be 2 × 2 and will have *T* and *T* along the top and *T* and *t* along the left side. Clockwise from the top left, the genotypes listed within the boxes will be *Tt*, *Tt*, *tt*, and *tt*. The phenotypic ratio will be 1 tall:1 dwarf.

Exercise:

Problem:

Use a Punnett square to predict the offspring in a cross between a tall pea plant (heterozygous) and a tall pea plant (heterozygous). What is the genotypic ratio of the offspring?

Solution:

The Punnett square will be 2 × 2 and will have *T* and *t* along the top and *T* and *t* along the left side. Clockwise from the top left, the genotypes listed within the boxes will be *TT*, *Tt*, *Tt*, and *tt*. The genotypic ratio will be 1*TT*:2*Tt*:1*tt*.

Glossary

allele
: one of two or more variants of a gene that determines a particular trait for a characteristic

dihybrid
: the result of a cross between two true-breeding parents that express different traits for two characteristics

genotype
: the underlying genetic makeup, consisting of both physically visible and non-expressed alleles, of an organism

heterozygous
: having two different alleles for a given gene on the homologous chromosomes

homozygous
: having two identical alleles for a given gene on the homologous chromosomes

law of dominance
: in a heterozygote, one trait will conceal the presence of another trait for the same characteristic

law of independent assortment
: genes do not influence each other with regard to sorting of alleles into gametes; every possible combination of alleles is equally likely to occur

law of segregation
: paired unit factors (i.e., genes) segregate equally into gametes such that offspring have an equal likelihood of inheriting any combination of factors

monohybrid

the result of a cross between two true-breeding parents that express different traits for only one characteristic

phenotype
the observable traits expressed by an organism

Punnett square
a visual representation of a cross between two individuals in which the gametes of each individual are denoted along the top and side of a grid, respectively, and the possible zygotic genotypes are recombined at each box in the grid

test cross
a cross between a dominant expressing individual with an unknown genotype and a homozygous recessive individual; the offspring phenotypes indicate whether the unknown parent is heterozygous or homozygous for the dominant trait

Extensions of the Laws of Inheritance
By the end of this section, you will be able to:

- Identify non-Mendelian inheritance patterns such as incomplete dominance, codominance, multiple alleles, and sex linkage from the results of crosses
- Explain the effect of linkage and recombination on gamete genotypes
- Explain the phenotypic outcomes of epistatic effects among genes

Mendel studied traits with only one mode of inheritance in pea plants. The inheritance of the traits he studied all followed the relatively simple pattern of dominant and recessive alleles for a single characteristic. There are several important modes of inheritance, discovered after Mendel's work, that do not follow the dominant and recessive, single-gene model.

Alternatives to Dominance and Recessiveness

Mendel's experiments with pea plants suggested that: 1) two types of "units" or alleles exist for every gene; 2) alleles maintain their integrity in each generation (no blending); and 3) in the presence of the dominant allele, the recessive allele is hidden, with no contribution to the phenotype. Therefore, recessive alleles can be "carried" and not expressed by individuals. Such heterozygous individuals are sometimes referred to as "carriers." Since then, genetic studies in other organisms have shown that much more complexity exists, but that the fundamental principles of Mendelian genetics still hold true. In the sections to follow, we consider some of the extensions of Mendelism.

Incomplete Dominance

Mendel's results, demonstrating that traits are inherited as dominant and recessive pairs, contradicted the view at that time that offspring exhibited a blend of their parents' traits. However, the heterozygote phenotype occasionally does appear to be intermediate between the two parents. For example, in the snapdragon, *Antirrhinum majus* ([link]), a cross between a homozygous parent with white flowers ($C^W C^W$) and a homozygous parent

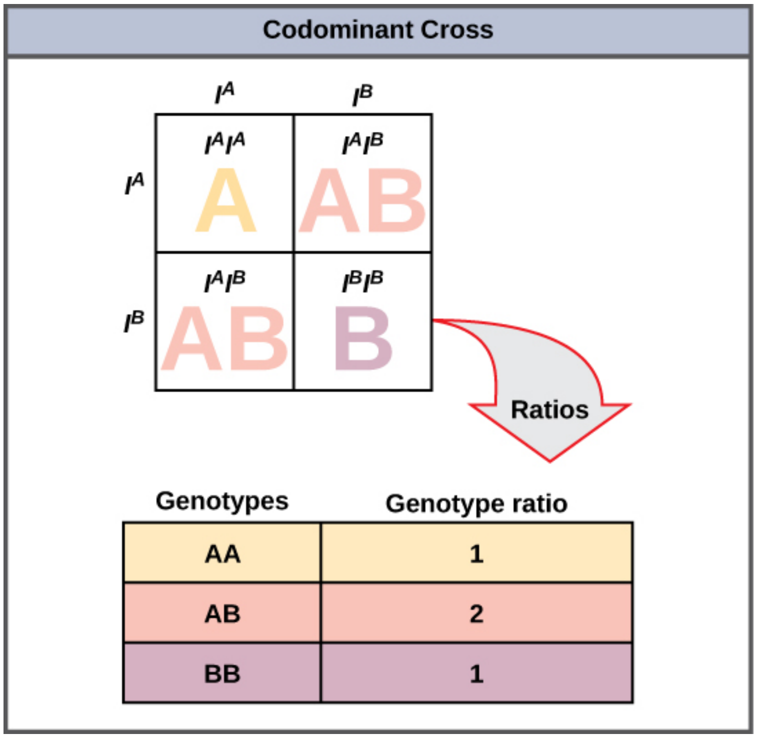
with red flowers (C^RC^R) will produce offspring with pink flowers (C^RC^W). (Note that different genotypic abbreviations are used for Mendelian extensions to distinguish these patterns from simple dominance and recessiveness.) This pattern of inheritance is described as **incomplete dominance**, meaning that one of the alleles appears in the phenotype in the heterozygote, but not to the exclusion of the other, which can also be seen. The allele for red flowers is incompletely dominant over the allele for white flowers. However, the results of a heterozygote self-cross can still be predicted, just as with Mendelian dominant and recessive crosses. In this case, the genotypic ratio would be 1 C^RC^R:2 C^RC^W:1 C^WC^W, and the phenotypic ratio would be 1:2:1 for red:pink:white. The basis for the intermediate color in the heterozygote is simply that the pigment produced by the red allele (anthocyanin) is diluted in the heterozygote and therefore appears pink because of the white background of the flower petals.

These pink flowers of a heterozygote snapdragon result from incomplete

dominance. (credit: "storebukkebruse"/Flickr)

Codominance

A variation on incomplete dominance is **codominance**, in which both alleles for the same characteristic are simultaneously expressed in the heterozygote. An example of codominance occurs in the ABO blood groups of humans. The A and B alleles are expressed in the form of A or B molecules present on the surface of red blood cells. Homozygotes (I^AI^A and I^BI^B) express either the A or the B phenotype, and heterozygotes (I^AI^B) express both phenotypes equally. The I^AI^B individual has blood type AB. In a self-cross between heterozygotes expressing a codominant trait, the three possible offspring genotypes are phenotypically distinct. However, the 1:2:1 genotypic ratio characteristic of a Mendelian monohybrid cross still applies ([link]).

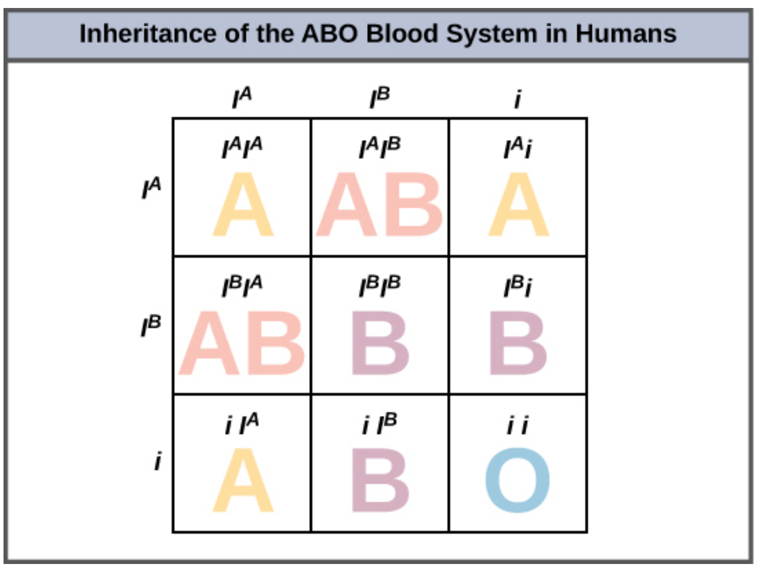


This Punnet square shows an AB/AB blood type cross

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level, such that many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype in the natural population as the **wild type** (often abbreviated "+"). All other phenotypes or genotypes are considered variants (mutants) of this typical form, meaning they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is the ABO blood-type system in humans. In this case, there are three alleles circulating in the population. The I^A allele codes for A molecules on the red blood cells, the I^B allele codes for B molecules on the surface of red blood cells, and the *i* allele codes for no molecules on the red blood cells. In this case, the I^A and I^B alleles are codominant with each other and are both dominant over the *i* allele. Although there are three alleles present in a population, each individual only gets two of the alleles from their parents. This produces the genotypes and phenotypes shown in [link]. Notice that instead of three genotypes, there are six different genotypes when there are three alleles. The number of possible phenotypes depends on the dominance relationships between the three alleles.



Inheritance of the ABO blood system in humans is shown.

Note:
Evolution in Action

Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae*, and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly. When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1 percent. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most effective malarial treatments can vary by geographic region.

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it is infective to humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait.

In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that comes about because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions in which this drug is widely used as an over-the-counter malaria remedy. As is common with pathogens that multiply to large numbers within an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide malaria burden.[footnote]

Sumiti Vinayak et al., "Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*," *PLoS Pathogens* 6 (2010): e1000830.

Sex-Linked Traits

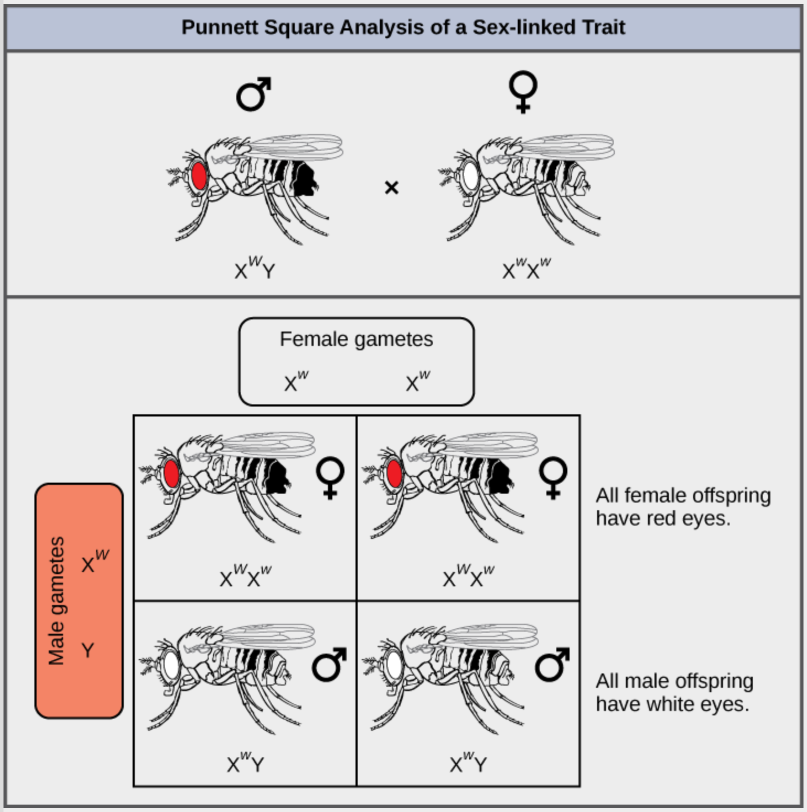
In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes—one pair of non-homologous chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or autosomes. In addition to 22 homologous pairs of autosomes, human females have a homologous pair of X chromosomes, whereas human males have an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains fewer genes. When a gene being examined is present on the X, but not the Y, chromosome, it is **X-linked**.

Eye color in *Drosophila*, the common fruit fly, was the first X-linked trait to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX. In flies the wild-type eye color is red (X^W) and is dominant to white eye color (X^w) ([link]). Because of the location of the eye-color gene, reciprocal crosses do not produce the same offspring ratios. Males are said to be **hemizygous**, in that they have only one allele for any X-linked characteristic. Hemizygosity makes descriptions of dominance and recessiveness irrelevant for XY males. *Drosophila* males lack the white gene on the Y chromosome; that is, their genotype can only be X^WY or X^wY. In contrast, females have two allele copies of this gene and can be X^WX^W, X^WX^w, or X^wX^w.

In *Drosophila*, the gene for eye color is located on the X chromosome. Red eye color is wild-type and is dominant to white eye color.

In an X-linked cross, the genotypes of F_1 and F_2 offspring depend on whether the recessive trait was expressed by the male or the female in the P generation. With respect to *Drosophila* eye color, when the P male expresses the white-eye phenotype and the female is homozygously red-eyed, all members of the F_1 generation exhibit red eyes ([link]). The F_1 females are heterozygous (X^WX^w), and the males are all X^WY, having received their X chromosome from the homozygous dominant P female and their Y chromosome from the P male. A subsequent cross between the X^WX^w female and the X^WY male would produce only red-eyed females (with X^WX^W or X^WX^w genotypes) and both red- and white-eyed males (with X^WY or X^wY genotypes). Now, consider a cross between a homozygous white-eyed female and a male with red eyes. The F_1 generation would exhibit only heterozygous red-eyed females (X^WX^w) and only white-eyed males (X^wY). Half of the F_2 females would be red-eyed (X^WX^w) and half would be white-eyed (X^wX^w). Similarly, half of the F_2 males would be red-eyed (X^WY) and half would be white-eyed (X^wY).

Note:
Art Connection



Crosses involving sex-linked traits often give rise to different phenotypes for the different sexes of offspring, as is the case for this cross involving red and white eye color in *Drosophila*. In the diagram, *w* is the white-eye mutant allele and *W* is the wild-type, red-eye allele.

What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

Discoveries in fruit fly genetics can be applied to human genetics. When a female parent is homozygous for a recessive X-linked trait, she will pass the

trait on to 100 percent of her male offspring, because the males will receive the Y chromosome from the male parent. In humans, the alleles for certain conditions (some color-blindness, hemophilia, and muscular dystrophy) are X-linked. Females who are heterozygous for these diseases are said to be carriers and may not exhibit any phenotypic effects. These females will pass the disease to half of their sons and will pass carrier status to half of their daughters; therefore, X-linked traits appear more frequently in males than females.

In some groups of organisms with sex chromosomes, the sex with the non-homologous sex chromosomes is the female rather than the male. This is the case for all birds. In this case, sex-linked traits will be more likely to appear in the female, in whom they are hemizygous.

Note:
Concept in Action

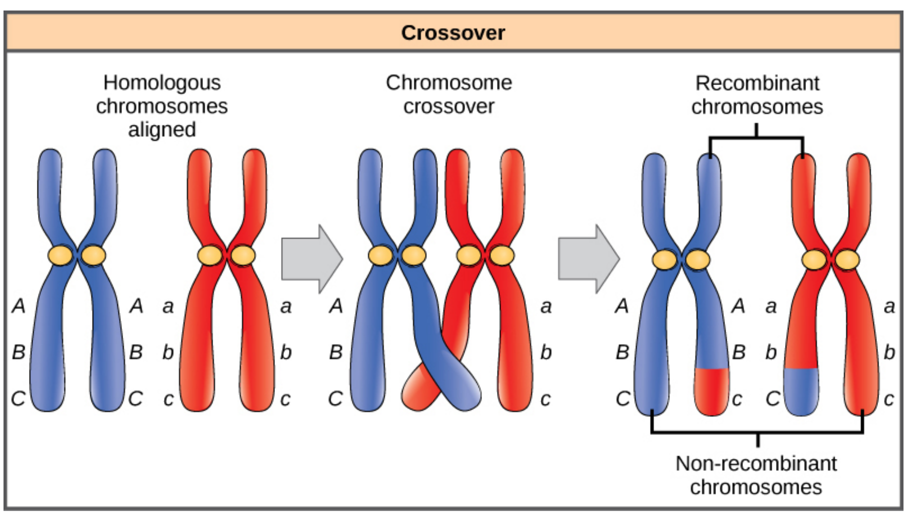
Watch this video to learn more about sex-linked traits.

Linked Genes Violate the Law of Independent Assortment

Although all of Mendel's pea plant characteristics behaved according to the law of independent assortment, we now know that some allele combinations are not inherited independently of each other. Genes that are located on separate, non-homologous chromosomes will always sort independently. However, each chromosome contains hundreds or thousands of genes, organized linearly on chromosomes like beads on a string. The segregation of alleles into gametes can be influenced by **linkage**, in which genes that

are located physically close to each other on the same chromosome are more likely to be inherited as a pair. However, because of the process of recombination, or “crossover,” it is possible for two genes on the same chromosome to behave independently, or as if they are not linked. To understand this, let us consider the biological basis of gene linkage and recombination.

Homologous chromosomes possess the same genes in the same order, though the specific alleles of the gene can be different on each of the two chromosomes. Recall that during interphase and prophase I of meiosis, homologous chromosomes first replicate and then synapse, with like genes on the homologs aligning with each other. At this stage, segments of homologous chromosomes exchange linear segments of genetic material ([link]). This process is called **recombination**, or crossover, and it is a common genetic process. Because the genes are aligned during recombination, the gene order is not altered. Instead, the result of recombination is that maternal and paternal alleles are combined onto the same chromosome. Across a given chromosome, several recombination events may occur, causing extensive shuffling of alleles.



The process of crossover, or recombination, occurs when two homologous chromosomes align and exchange a segment of genetic material.

When two genes are located on the same chromosome, they are considered linked, and their alleles tend to be transmitted through meiosis together. To exemplify this, imagine a dihybrid cross involving flower color and plant height in which the genes are next to each other on the chromosome. If one homologous chromosome has alleles for tall plants and red flowers, and the other chromosome has genes for short plants and yellow flowers, then when the gametes are formed, the tall and red alleles will tend to go together into a gamete and the short and yellow alleles will go into other gametes. These are called the parental genotypes because they have been inherited intact from the parents of the individual producing gametes. But unlike if the genes were on different chromosomes, there will be no gametes with tall and yellow alleles and no gametes with short and red alleles. If you create a Punnett square with these gametes, you will see that the classical Mendelian prediction of a 9:3:3:1 outcome of a dihybrid cross would not apply. As the distance between two genes increases, the probability of one or more crossovers between them increases and the genes behave more like they are on separate chromosomes. Geneticists have used the proportion of recombinant gametes (the ones not like the parents) as a measure of how far apart genes are on a chromosome. Using this information, they have constructed linkage maps of genes on chromosomes for well-studied organisms, including humans.

Mendel’s seminal publication makes no mention of linkage, and many researchers have questioned whether he encountered linkage but chose not to publish those crosses out of concern that they would invalidate his independent assortment postulate. The garden pea has seven chromosomes, and some have suggested that his choice of seven characteristics was not a coincidence. However, even if the genes he examined were not located on separate chromosomes, it is possible that he simply did not observe linkage because of the extensive shuffling effects of recombination.

Epistasis

Mendel’s studies in pea plants implied that the sum of an individual’s phenotype was controlled by genes (or as he called them, unit factors), such

that every characteristic was distinctly and completely controlled by a single gene. In fact, single observable characteristics are almost always under the influence of multiple genes (each with two or more alleles) acting in unison. For example, at least eight genes contribute to eye color in humans.

Note:
Concept in Action

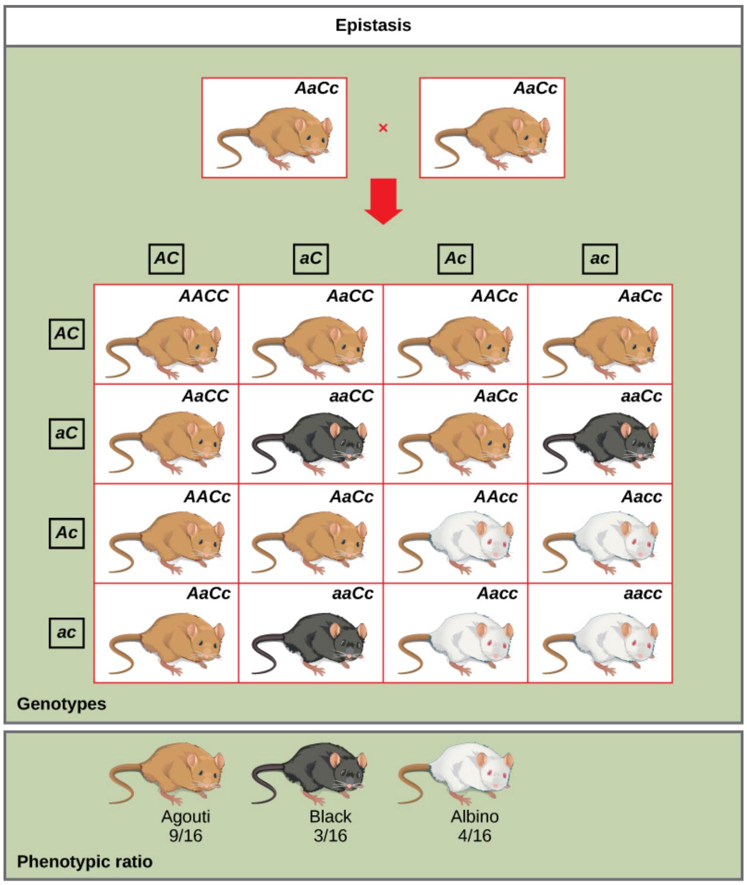
Eye color in humans is determined by multiple alleles. Use the Eye Color Calculator to predict the eye color of children from parental eye color.

In some cases, several genes can contribute to aspects of a common phenotype without their gene products ever directly interacting. In the case of organ development, for instance, genes may be expressed sequentially, with each gene adding to the complexity and specificity of the organ. Genes may function in complementary or synergistic fashions, such that two or more genes expressed simultaneously affect a phenotype. An apparent example of this occurs with human skin color, which appears to involve the action of at least three (and probably more) genes. Cases in which inheritance for a characteristic like skin color or human height depend on the combined effects of numerous genes are called polygenic inheritance.

Genes may also oppose each other, with one gene suppressing the expression of another. In **epistasis**, the interaction between genes is antagonistic, such that one gene masks or interferes with the expression of another. “Epistasis” is a word composed of Greek roots meaning “standing upon.” The alleles that are being masked or silenced are said to be

hypostatic to the epistatic alleles that are doing the masking. Often the biochemical basis of epistasis is a gene pathway in which expression of one gene is dependent on the function of a gene that precedes or follows it in the pathway.

An example of epistasis is pigmentation in mice. The wild-type coat color, agouti (AA) is dominant to solid-colored fur (aa). However, a separate gene C, when present as the recessive homozygote (cc), negates any expression of pigment from the A gene and results in an albino mouse ([link]). Therefore, the genotypes *AAcc*, *Aacc*, and *aacc* all produce the same albino phenotype. A cross between heterozygotes for both genes (*AaCc* x *AaCc*) would generate offspring with a phenotypic ratio of 9 agouti:3 black:4 albino ([link]). In this case, the *C* gene is epistatic to the *A* gene.



In this example of epistasis, one gene

(*C*) masks the expression of another (*A*) for coat color. When the *C* allele is present, coat color is expressed; when it is absent (*cc*), no coat color is expressed. Coat color depends on the *A* gene, which shows dominance, with the recessive homozygote showing a different phenotype than the heterozygote or dominant homozygote.

Section Summary

Alleles do not always behave in dominant and recessive patterns. Incomplete dominance describes situations in which the heterozygote exhibits a phenotype that is intermediate between the homozygous phenotypes. Codominance describes the simultaneous expression of both of the alleles in the heterozygote. Although diploid organisms can only have two alleles for any given gene, it is common for more than two alleles for a gene to exist in a population. In humans, as in many animals and some plants, females have two X chromosomes and males have one X and one Y chromosome. Genes that are present on the X but not the Y chromosome are said to be X-linked, such that males only inherit one allele for the gene, and females inherit two.

According to Mendel’s law of independent assortment, genes sort independently of each other into gametes during meiosis. This occurs because chromosomes, on which the genes reside, assort independently during meiosis and crossovers cause most genes on the same chromosomes to also behave independently. When genes are located in close proximity on the same chromosome, their alleles tend to be inherited together. This results in offspring ratios that violate Mendel's law of independent assortment. However, recombination serves to exchange genetic material on homologous chromosomes such that maternal and paternal alleles may be recombined on the same chromosome. This is why alleles on a given chromosome are not always inherited together. Recombination is a random

event occurring anywhere on a chromosome. Therefore, genes that are far apart on the same chromosome are likely to still assort independently because of recombination events that occurred in the intervening chromosomal space.

Whether or not they are sorting independently, genes may interact at the level of gene products, such that the expression of an allele for one gene masks or modifies the expression of an allele for a different gene. This is called epistasis.

Art Connections

Exercise:

Problem:

[link] What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

Solution:

[link] Half of the female offspring would be heterozygous (X^WX^w) with red eyes, and half would be homozygous recessive (X^wX^w) with white eyes. Half of the male offspring would be hemizygous dominant (X^WY) with red eyes, and half would be hemizygous recessive (X^wY) with white eyes.

Multiple Choice

Exercise:

Problem:

If black and white true-breeding mice are mated and the result is all gray offspring, what inheritance pattern would this be indicative of?

a. dominance
b. codominance

c. multiple alleles
d. incomplete dominance

Solution:

D

Exercise:

Problem:

The ABO blood groups in humans are expressed as the I^A, I^B, and i alleles. The I^A allele encodes the A blood group antigen, I^B encodes B, and i encodes O. Both A and B are dominant to O. If a heterozygous blood type A parent (I^Ai) and a heterozygous blood type B parent (I^Bi) mate, one quarter of their offspring are expected to have the AB blood type (I^AI^B) in which both antigens are expressed equally. Therefore, ABO blood groups are an example of:

a. multiple alleles and incomplete dominance
b. codominance and incomplete dominance
c. incomplete dominance only
d. multiple alleles and codominance

Solution:

D

Exercise:

Problem:

In a cross between a homozygous red-eyed female fruit fly and a white-eyed male fruit fly, what is the expected outcome?

a. all white-eyed male offspring
b. all white-eyed female offspring
c. all red-eyed offspring

d. half white-eyed make offspring

Solution:

C

Exercise:

Problem:

When a population has a gene with four alleles circulating, how many possible genotypes are there?

a. 3
b. 6
c. 10
d. 16

Solution:

C

Free Response

Exercise:

Problem: Can a male be a carrier of red-green color blindness?

Solution:

No, males can only express color blindness and cannot carry it because an individual needs two X chromosomes to be a carrier.

Exercise:

Problem:

Could an individual with blood type O (genotype *ii*) be a legitimate child of parents in which one parent had blood type A and the other parent had blood type B?

Solution:

Yes this child could have come from these parents. The child would have inherited an *i* allele from each parent and for this to happen the type A parent had to have genotype $I^A i$ and the type b parent had to have genotype $I^B i$.

Glossary

codominance
: in a heterozygote, complete and simultaneous expression of both alleles for the same characteristic

epistasis
: an interaction between genes such that one gene masks or interferes with the expression of another

hemizygous
: the presence of only one allele for a characteristic, as in X-linkage; hemizygosity makes descriptions of dominance and recessiveness irrelevant

incomplete dominance
: in a heterozygote, expression of two contrasting alleles such that the individual displays an intermediate phenotype

linkage
: a phenomenon in which alleles that are located in close proximity to each other on the same chromosome are more likely to be inherited together

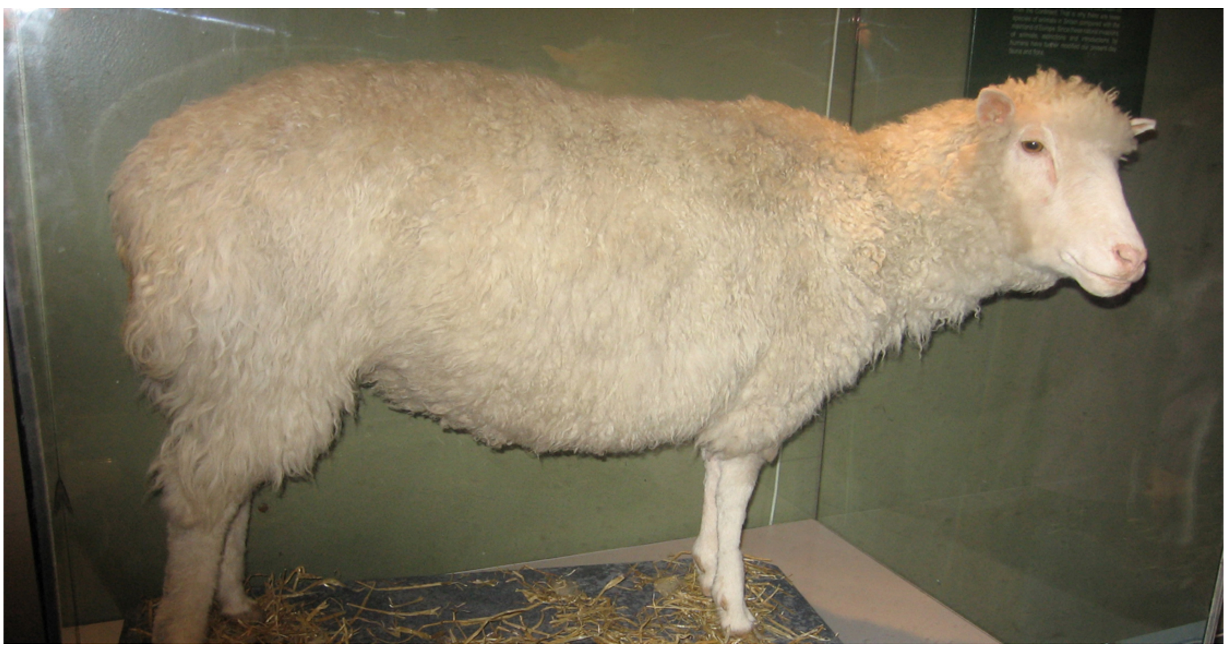
recombination
: the process during meiosis in which homologous chromosomes exchange linear segments of genetic material, thereby dramatically increasing genetic variation in the offspring and separating linked genes

wild type
: the most commonly occurring genotype or phenotype for a given characteristic found in a population

X-linked
: a gene present on the X chromosome, but not the Y chromosome

Introduction
class="introduction"

Dolly the sheep was the first cloned mammal.



The three letters “DNA” have now become associated with crime solving, paternity testing, human identification, and genetic testing. DNA can be retrieved from hair, blood, or saliva. With the exception of identical twins, each person’s DNA is unique and it is possible to detect differences between human beings on the basis of their unique DNA sequence.

DNA analysis has many practical applications beyond forensics and paternity testing. DNA testing is used for tracing genealogy and identifying pathogens. In the medical field, DNA is used in diagnostics, new vaccine

development, and cancer therapy. It is now possible to determine predisposition to many diseases by analyzing genes.

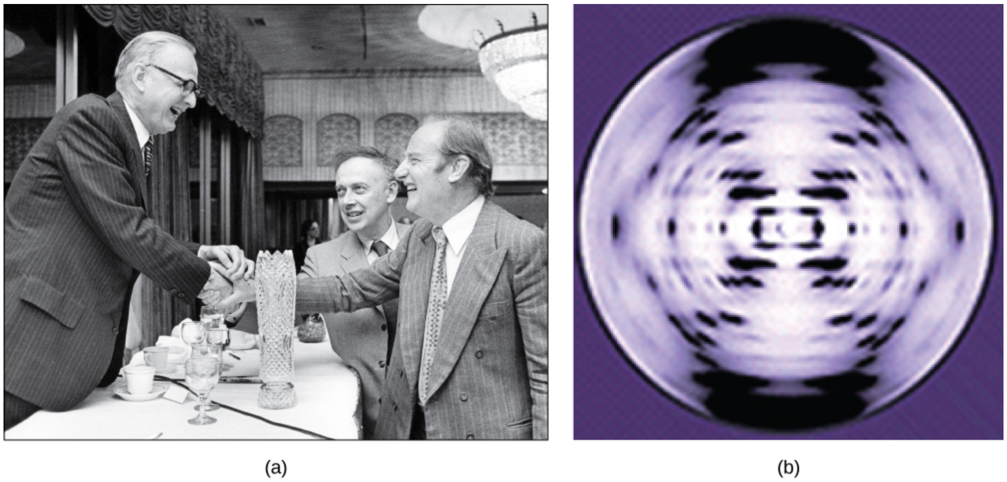
DNA is the genetic material passed from parent to offspring for all life on Earth. The technology of molecular genetics developed in the last half century has enabled us to see deep into the history of life to deduce the relationships between living things in ways never thought possible. It also allows us to understand the workings of evolution in populations of organisms. Over a thousand species have had their entire genome sequenced, and there have been thousands of individual human genome sequences completed. These sequences will allow us to understand human disease and the relationship of humans to the rest of the tree of life. Finally, molecular genetics techniques have revolutionized plant and animal breeding for human agricultural needs. All of these advances in biotechnology depended on basic research leading to the discovery of the structure of DNA in 1953, and the research since then that has uncovered the details of DNA replication and the complex process leading to the expression of DNA in the form of proteins in the cell.

The Structure of DNA

By the end of this section, you will be able to:

- Describe the structure of DNA
- Describe how eukaryotic and prokaryotic DNA is arranged in the cell

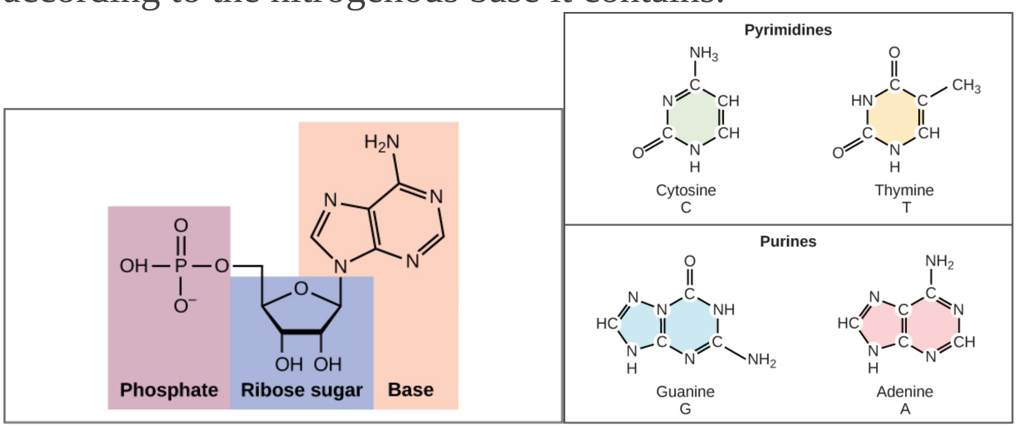
In the 1950s, Francis Crick and James Watson worked together at the University of Cambridge, England, to determine the structure of DNA. Other scientists, such as Linus Pauling and Maurice Wilkins, were also actively exploring this field. Pauling had discovered the secondary structure of proteins using X-ray crystallography. X-ray crystallography is a method for investigating molecular structure by observing the patterns formed by X-rays shot through a crystal of the substance. The patterns give important information about the structure of the molecule of interest. In Wilkins' lab, researcher Rosalind Franklin was using X-ray crystallography to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule using Franklin's data ([link]). Watson and Crick also had key pieces of information available from other researchers such as Chargaff's rules. Chargaff had shown that of the four kinds of monomers (nucleotides) present in a DNA molecule, two types were always present in equal amounts and the remaining two types were also always present in equal amounts. This meant they were always paired in some way. In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine for their work in determining the structure of DNA.



Pioneering scientists (a) James Watson and Francis Crick

are pictured here with American geneticist Maclyn McCarty. Scientist Rosalind Franklin discovered (b) the X-ray diffraction pattern of DNA, which helped to elucidate its double helix structure. (credit a: modification of work by Marjorie McCarty; b: modification of work by NIH)

Now let’s consider the structure of the two types of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The building blocks of DNA are nucleotides, which are made up of three parts: a **deoxyribose** (5-carbon sugar), a **phosphate group**, and a **nitrogenous base** ([link]). There are four types of nitrogenous bases in DNA. Adenine (A) and guanine (G) are double-ringed purines, and cytosine (C) and thymine (T) are smaller, single-ringed pyrimidines. The nucleotide is named according to the nitrogenous base it contains.

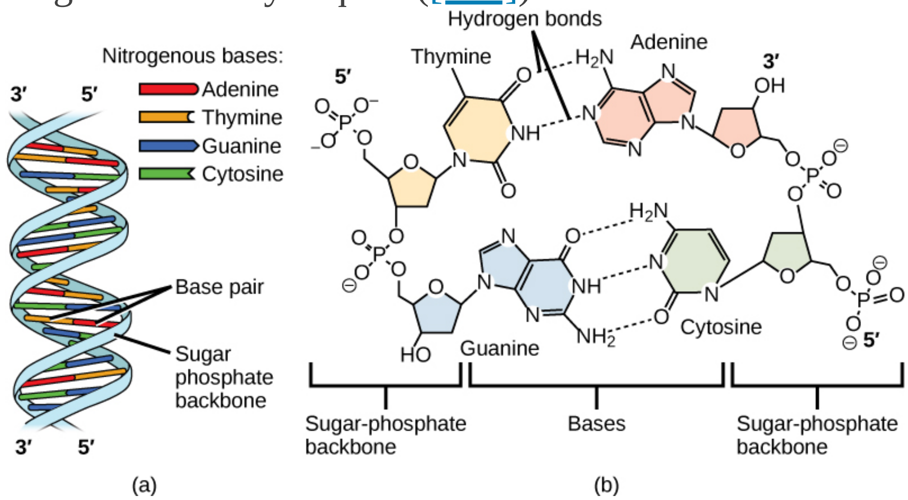


(a) Each DNA nucleotide is made up of a sugar, a phosphate group, and a base. (b) Cytosine and thymine are pyrimidines. Guanine and adenine are purines.

The phosphate group of one nucleotide bonds covalently with the sugar molecule of the next nucleotide, and so on, forming a long polymer of nucleotide monomers. The sugar–phosphate groups line up in a “backbone” for each single strand of DNA, and the nucleotide bases stick out from this

backbone. The carbon atoms of the five-carbon sugar are numbered clockwise from the oxygen as 1', 2', 3', 4', and 5' (1' is read as “one prime”). The phosphate group is attached to the 5' carbon of one nucleotide and the 3' carbon of the next nucleotide. In its natural state, each DNA molecule is actually composed of two single strands held together along their length with hydrogen bonds between the bases.

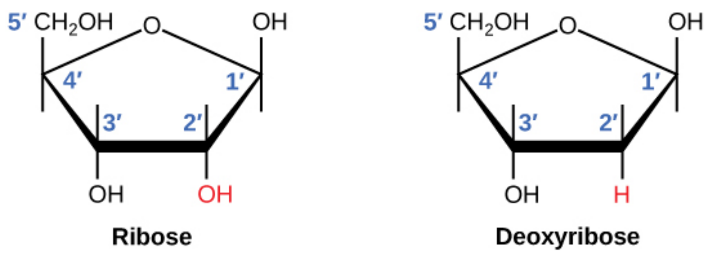
Watson and Crick proposed that the DNA is made up of two strands that are twisted around each other to form a right-handed helix, called a **double helix**. Base-pairing takes place between a purine and pyrimidine: namely, A pairs with T, and G pairs with C. In other words, adenine and thymine are complementary base pairs, and cytosine and guanine are also complementary base pairs. This is the basis for Chargaff’s rule; because of their complementarity, there is as much adenine as thymine in a DNA molecule and as much guanine as cytosine. Adenine and thymine are connected by two hydrogen bonds, and cytosine and guanine are connected by three hydrogen bonds. The two strands are anti-parallel in nature; that is, one strand will have the 3' carbon of the sugar in the “upward” position, whereas the other strand will have the 5' carbon in the upward position. The diameter of the DNA double helix is uniform throughout because a purine (two rings) always pairs with a pyrimidine (one ring) and their combined lengths are always equal. ([link]).



DNA (a) forms a double stranded helix, and (b) adenine pairs with thymine and cytosine pairs with guanine. (credit a: modification of work by Jerome Walker, Dennis Myts)

The Structure of RNA

There is a second nucleic acid in all cells called ribonucleic acid, or RNA. Like DNA, RNA is a polymer of nucleotides. Each of the nucleotides in RNA is made up of a nitrogenous base, a five-carbon sugar, and a phosphate group. In the case of RNA, the five-carbon sugar is ribose, not deoxyribose. Ribose has a hydroxyl group at the 2' carbon, unlike deoxyribose, which has only a hydrogen atom ([link]).

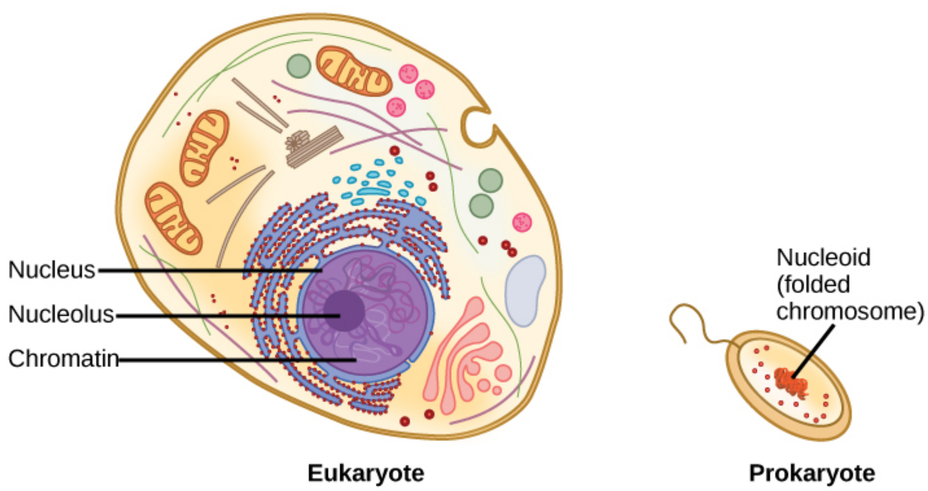


The difference between the ribose found in RNA and the deoxyribose found in DNA is that ribose has a hydroxyl group at the 2' carbon.

RNA nucleotides contain the nitrogenous bases adenine, cytosine, and guanine. However, they do not contain thymine, which is instead replaced by uracil, symbolized by a “U.” RNA exists as a single-stranded molecule rather than a double-stranded helix. Molecular biologists have named several kinds of RNA on the basis of their function. These include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA)—molecules that are involved in the production of proteins from the DNA code.

How DNA Is Arranged in the Cell

DNA is a working molecule; it must be replicated when a cell is ready to divide, and it must be “read” to produce the molecules, such as proteins, to carry out the functions of the cell. For this reason, the DNA is protected and packaged in very specific ways. In addition, DNA molecules can be very long. Stretched end-to-end, the DNA molecules in a single human cell would come to a length of about 2 meters. Thus, the DNA for a cell must be packaged in a very ordered way to fit and function within a structure (the cell) that is not visible to the naked eye. The chromosomes of prokaryotes are much simpler than those of eukaryotes in many of their features ([link]). Most prokaryotes contain a single, circular chromosome that is found in an area in the cytoplasm called the nucleoid.



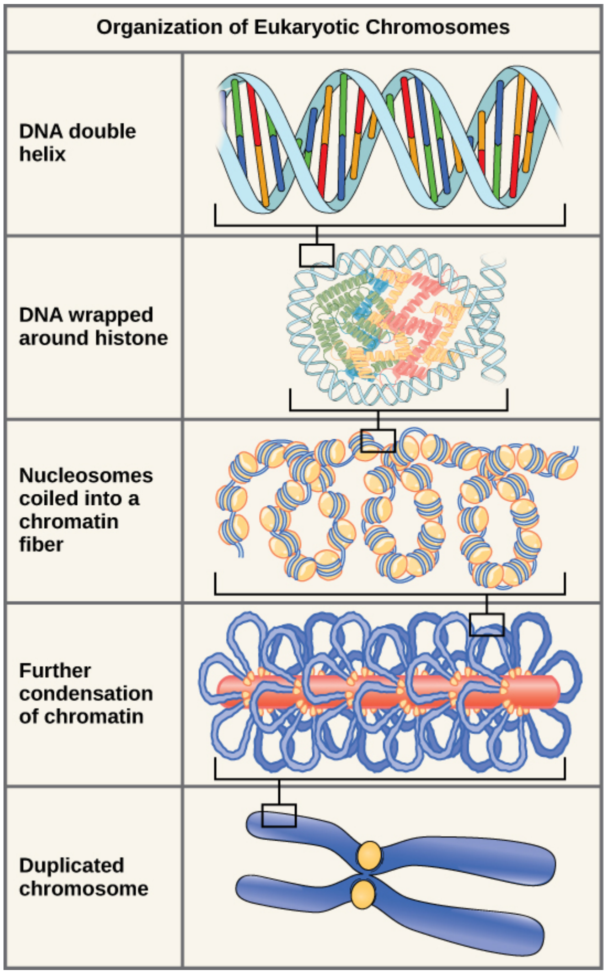
A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

The size of the genome in one of the most well-studied prokaryotes, *Escherichia coli,* is 4.6 million base pairs, which would extend a distance of about 1.6 mm if stretched out. So how does this fit inside a small bacterial cell? The DNA is twisted beyond the double helix in what is known as

supercoiling. Some proteins are known to be involved in the supercoiling; other proteins and enzymes help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus ([link]). At the most basic level, DNA is wrapped around proteins known as histones to form structures called nucleosomes. The DNA is wrapped tightly around the histone core. This nucleosome is linked to the next one by a short strand of DNA that is free of histones. This is also known as the “beads on a string” structure; the nucleosomes are the “beads” and the short lengths of DNA between them are the “string.” The nucleosomes, with their DNA coiled around them, stack compactly onto each other to form a 30-nm–wide fiber. This fiber is further coiled into a thicker and more compact structure. At the metaphase stage of mitosis, when the chromosomes are lined up in the center of the cell, the chromosomes are at their most compacted. They are approximately 700 nm in width, and are found in association with scaffold proteins.

In interphase, the phase of the cell cycle between mitoses at which the chromosomes are decondensed, eukaryotic chromosomes have two distinct regions that can be distinguished by staining. There is a tightly packaged region that stains darkly, and a less dense region. The darkly staining regions usually contain genes that are not active, and are found in the regions of the centromere and telomeres. The lightly staining regions usually contain genes that are active, with DNA packaged around nucleosomes but not further compacted.



These figures illustrate the compaction of the eukaryotic chromosome.

Note:
Concept in Action

Watch this animation of DNA packaging.

Section Summary

The model of the double-helix structure of DNA was proposed by Watson and Crick. The DNA molecule is a polymer of nucleotides. Each nucleotide is composed of a nitrogenous base, a five-carbon sugar (deoxyribose), and a phosphate group. There are four nitrogenous bases in DNA, two purines (adenine and guanine) and two pyrimidines (cytosine and thymine). A DNA molecule is composed of two strands. Each strand is composed of nucleotides bonded together covalently between the phosphate group of one and the deoxyribose sugar of the next. From this backbone extend the bases. The bases of one strand bond to the bases of the second strand with hydrogen bonds. Adenine always bonds with thymine, and cytosine always bonds with guanine. The bonding causes the two strands to spiral around each other in a shape called a double helix. Ribonucleic acid (RNA) is a second nucleic acid found in cells. RNA is a single-stranded polymer of nucleotides. It also differs from DNA in that it contains the sugar ribose, rather than deoxyribose, and the nucleotide uracil rather than thymine. Various RNA molecules function in the process of forming proteins from the genetic code in DNA.

Prokaryotes contain a single, double-stranded circular chromosome. Eukaryotes contain double-stranded linear DNA molecules packaged into chromosomes. The DNA helix is wrapped around proteins to form nucleosomes. The protein coils are further coiled, and during mitosis and meiosis, the chromosomes become even more greatly coiled to facilitate their movement. Chromosomes have two distinct regions which can be distinguished by staining, reflecting different degrees of packaging and

determined by whether the DNA in a region is being expressed (euchromatin) or not (heterochromatin).

Multiple Choice

Exercise:

Problem: Which of the following does cytosine pair with?

a. guanine
b. thymine
c. adenine
d. a pyrimidine

Solution:

A

Exercise:

Problem:

Prokaryotes contain a _________chromosome, and eukaryotes contain _________ chromosomes.

a. single-stranded circular; single-stranded linear
b. single-stranded linear; single-stranded circular
c. double-stranded circular; double-stranded linear
d. double-stranded linear; double-stranded circular

Solution:

C

Free Response

Exercise:

Problem: Describe the organization of the eukaryotic chromosome.

Solution:

The DNA is wound around proteins called histones. The histones then stack together in a compact form that creates a fiber that is 30-nm thick. The fiber is further coiled for greater compactness. During metaphase of mitosis, the chromosome is at its most compact to facilitate chromosome movement. During interphase, there are denser areas of chromatin, called heterochromatin, that contain DNA that is not expressed, and less dense euchromatin that contains DNA that is expressed.

Exercise:

Problem:

Describe the structure and complementary base pairing of DNA.

Solution:

A single strand of DNA is a polymer of nucleic acids joined covalently between the phosphate group of one and the deoxyribose sugar of the next to for a “backbone” from which the nitrogenous bases stick out. In its natural state, DNA has two strands wound around each other in a double helix. The bases on each strand are bonded to each other with hydrogen bonds. Only specific bases bond with each other; adenine bonds with thymine, and cytosine bonds with guanine.

Glossary

deoxyribose
: a five-carbon sugar molecule with a hydrogen atom rather than a hydroxyl group in the 2' position; the sugar component of DNA nucleotides

double helix

the molecular shape of DNA in which two strands of nucleotides wind around each other in a spiral shape

nitrogenous base

a nitrogen-containing molecule that acts as a base; often referring to one of the purine or pyrimidine components of nucleic acids

phosphate group

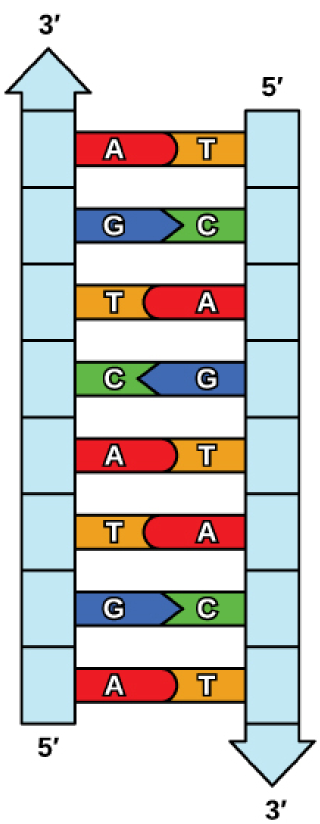
a molecular group consisting of a central phosphorus atom bound to four oxygen atoms

DNA Replication
By the end of this section, you will be able to:

- Explain the process of DNA replication
- Explain the importance of telomerase to DNA replication
- Describe mechanisms of DNA repair

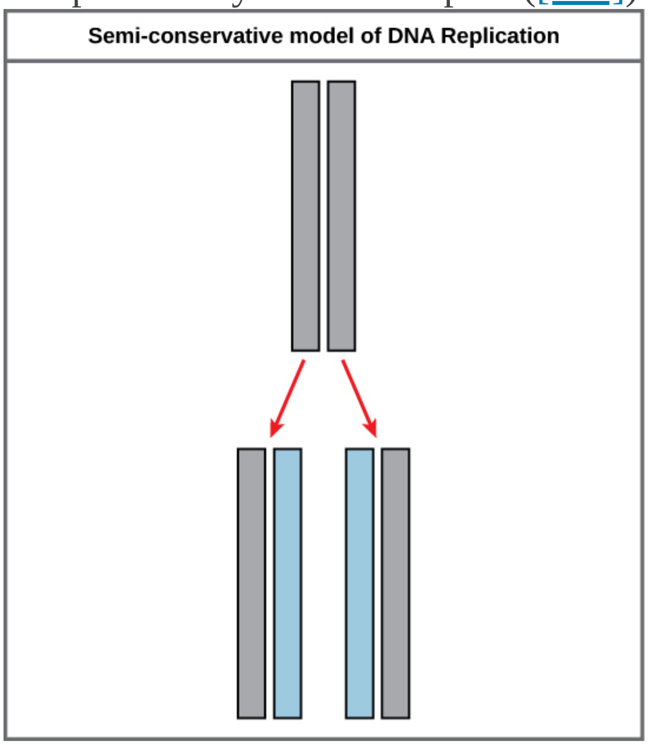
When a cell divides, it is important that each daughter cell receives an identical copy of the DNA. This is accomplished by the process of DNA replication. The replication of DNA occurs during the synthesis phase, or S phase, of the cell cycle, before the cell enters mitosis or meiosis.

The elucidation of the structure of the double helix provided a hint as to how DNA is copied. Recall that adenine nucleotides pair with thymine nucleotides, and cytosine with guanine. This means that the two strands are complementary to each other. For example, a strand of DNA with a nucleotide sequence of AGTCATGA will have a complementary strand with the sequence TCAGTACT ([link]).



The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

Because of the complementarity of the two strands, having one strand means that it is possible to recreate the other strand. This model for replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied ([link]).



The semiconservative model of DNA replication is shown. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

During DNA replication, each of the two strands that make up the double helix serves as a template from which new strands are copied. The new strand will be complementary to the parental or “old” strand. Each new double strand consists of one parental strand and one new daughter strand. This is known as **semiconservative replication**. When two DNA copies are formed, they have an identical sequence of nucleotide bases and are divided equally into two daughter cells.

DNA Replication in Eukaryotes

Because eukaryotic genomes are very complex, DNA replication is a very complicated process that involves several enzymes and other proteins. It occurs in three main stages: initiation, elongation, and termination.

Recall that eukaryotic DNA is bound to proteins known as histones to form structures called nucleosomes. During initiation, the DNA is made accessible to the proteins and enzymes involved in the replication process. How does the replication machinery know where on the DNA double helix to begin? It turns out that there are specific nucleotide sequences called origins of replication at which replication begins. Certain proteins bind to the origin of replication while an enzyme called **helicase** unwinds and opens up the DNA helix. As the DNA opens up, Y-shaped structures called **replication forks** are formed ([link]). Two replication forks are formed at the origin of replication, and these get extended in both directions as replication proceeds. There are multiple origins of replication on the eukaryotic chromosome, such that replication can occur simultaneously from several places in the genome.

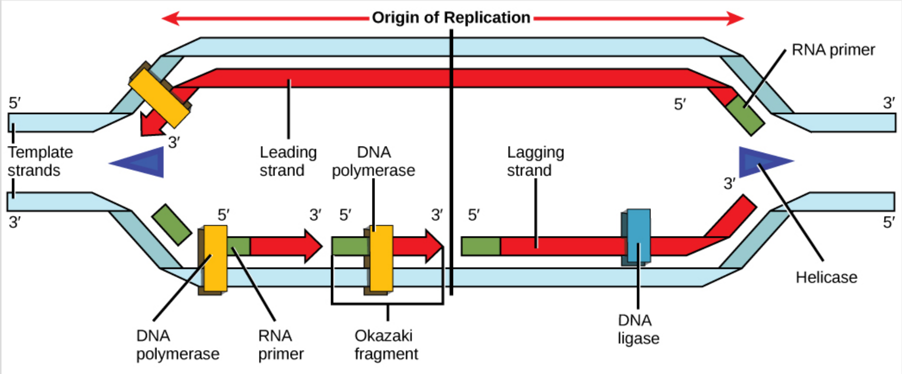
During elongation, an enzyme called **DNA polymerase** adds DNA nucleotides to the 3' end of the template. Because DNA polymerase can only add new nucleotides at the end of a backbone, a **primer** sequence, which provides this starting point, is added with complementary RNA nucleotides. This primer is removed later, and the nucleotides are replaced with DNA nucleotides. One strand, which is complementary to the parental DNA strand, is synthesized continuously toward the replication fork so the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. Because DNA

polymerase can only synthesize DNA in a 5' to 3' direction, the other new strand is put together in short pieces called **Okazaki fragments**. The Okazaki fragments each require a primer made of RNA to start the synthesis. The strand with the Okazaki fragments is known as the **lagging strand**. As synthesis proceeds, an enzyme removes the RNA primer, which is then replaced with DNA nucleotides, and the gaps between fragments are sealed by an enzyme called **DNA ligase**.

The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.
2. New bases are added to the complementary parental strands. One new strand is made continuously, while the other strand is made in pieces.
3. Primers are removed, new DNA nucleotides are put in place of the primers and the backbone is sealed by DNA ligase.

Note:
Art Connection

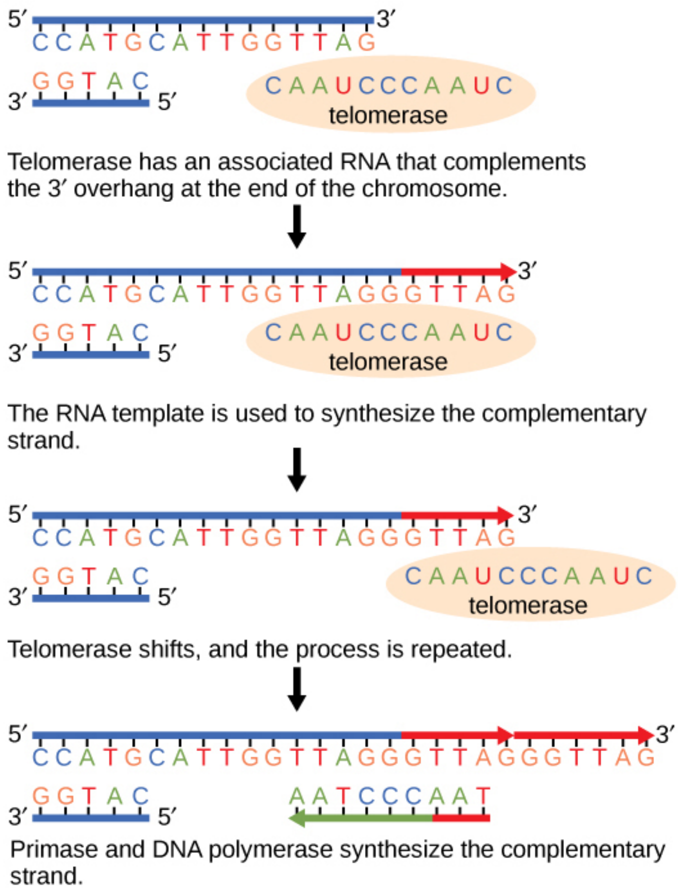


A replication fork is formed by the opening of the origin of replication, and helicase separates the DNA strands. An RNA primer is synthesized, and is elongated by the DNA polymerase. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches. The DNA fragments are joined by DNA ligase (not shown).

You isolate a cell strain in which the joining together of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

Telomere Replication

Because eukaryotic chromosomes are linear, DNA replication comes to the end of a line in eukaryotic chromosomes. As you have learned, the DNA polymerase enzyme can add nucleotides in only one direction. In the leading strand, synthesis continues until the end of the chromosome is reached; however, on the lagging strand there is no place for a primer to be made for the DNA fragment to be copied at the end of the chromosome. This presents a problem for the cell because the ends remain unpaired, and over time these ends get progressively shorter as cells continue to divide. The ends of the linear chromosomes are known as **telomeres**, which have repetitive sequences that do not code for a particular gene. As a consequence, it is telomeres that are shortened with each round of DNA replication instead of genes. For example, in humans, a six base-pair sequence, TTAGGG, is repeated 100 to 1000 times. The discovery of the enzyme **telomerase** ([link]) helped in the understanding of how chromosome ends are maintained. The telomerase attaches to the end of the chromosome, and complementary bases to the RNA template are added on the end of the DNA strand. Once the lagging strand template is sufficiently elongated, DNA polymerase can now add nucleotides that are complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.



The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

Telomerase is typically found to be active in germ cells, adult stem cells, and some cancer cells. For her discovery of telomerase and its action, Elizabeth Blackburn ([link]) received the Nobel Prize for Medicine and Physiology in 2009.

Elizabeth Blackburn, 2009 Nobel Laureate, was the scientist who discovered how telomerase works. (credit: U.S. Embassy, Stockholm, Sweden)

Telomerase is not active in adult somatic cells. Adult somatic cells that undergo cell division continue to have their telomeres shortened. This essentially means that telomere shortening is associated with aging. In 2010, scientists found that telomerase can reverse some age-related conditions in mice, and this may have potential in regenerative medicine. [footnote] Telomerase-deficient mice were used in these studies; these mice have tissue atrophy, stem-cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved functioning of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

Mariella Jaskelioff, et al., "Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice," *Nature*, 469 (2011):102–7.

DNA Replication in Prokaryotes

Recall that the prokaryotic chromosome is a circular molecule with a less extensive coiling structure than eukaryotic chromosomes. The eukaryotic chromosome is linear and highly coiled around proteins. While there are many similarities in the DNA replication process, these structural differences necessitate some differences in the DNA replication process in these two life forms.

DNA replication has been extremely well-studied in prokaryotes, primarily because of the small size of the genome and large number of variants available. *Escherichia coli* has 4.6 million base pairs in a single circular chromosome, and all of it gets replicated in approximately 42 minutes, starting from a single origin of replication and proceeding around the chromosome in both directions. This means that approximately 1000 nucleotides are added per second. The process is much more rapid than in eukaryotes. [link] summarizes the differences between prokaryotic and eukaryotic replications.

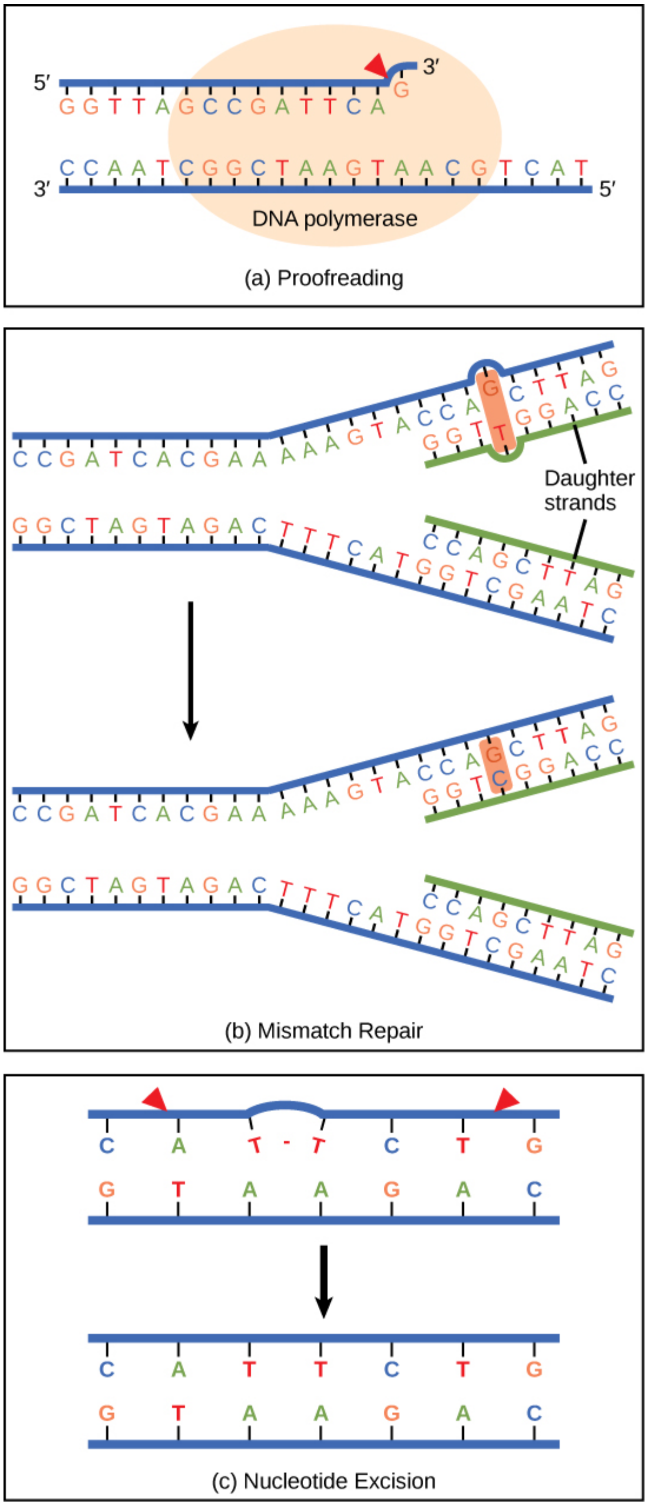
Differences between Prokaryotic and Eukaryotic Replications		
Property	**Prokaryotes**	**Eukaryotes**
Origin of replication	Single	Multiple
Rate of replication	1000 nucleotides/s	50 to 100 nucleotides/s
Chromosome structure	circular	linear
Telomerase	Not present	Present

Note:
Concept in Action

Click through a tutorial on DNA replication.

DNA Repair

DNA polymerase can make mistakes while adding nucleotides. It edits the DNA by proofreading every newly added base. Incorrect bases are removed and replaced by the correct base, and then polymerization continues ([link]**a**). Most mistakes are corrected during replication, although when this does not happen, the **mismatch repair** mechanism is employed. Mismatch repair enzymes recognize the wrongly incorporated base and excise it from the DNA, replacing it with the correct base ([link]**b**). In yet another type of repair, **nucleotide excision repair**, the DNA double strand is unwound and separated, the incorrect bases are removed along with a few bases on the 5' and 3' end, and these are replaced by copying the template with the help of DNA polymerase ([link]**c**). Nucleotide excision repair is particularly important in correcting thymine dimers, which are primarily caused by ultraviolet light. In a thymine dimer, two thymine nucleotides adjacent to each other on one strand are covalently bonded to each other rather than their complementary bases. If the dimer is not removed and repaired it will lead to a mutation. Individuals with flaws in their nucleotide excision repair genes show extreme sensitivity to sunlight and develop skin cancers early in life.



Proofreading by DNA polymerase (a) corrects errors during replication. In mismatch repair (b),

the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base. Nucleotide excision (c) repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.

Most mistakes are corrected; if they are not, they may result in a **mutation**—defined as a permanent change in the DNA sequence. Mutations in repair genes may lead to serious consequences like cancer.

Section Summary

DNA replicates by a semi-conservative method in which each of the two parental DNA strands act as a template for new DNA to be synthesized. After replication, each DNA has one parental or “old” strand, and one daughter or “new” strand.

Replication in eukaryotes starts at multiple origins of replication, while replication in prokaryotes starts from a single origin of replication. The DNA is opened with enzymes, resulting in the formation of the replication fork. Primase synthesizes an RNA primer to initiate synthesis by DNA polymerase, which can add nucleotides in only one direction. One strand is synthesized continuously in the direction of the replication fork; this is called the leading strand. The other strand is synthesized in a direction away from the replication fork, in short stretches of DNA known as Okazaki fragments. This strand is known as the lagging strand. Once replication is

completed, the RNA primers are replaced by DNA nucleotides and the DNA is sealed with DNA ligase.

The ends of eukaryotic chromosomes pose a problem, as polymerase is unable to extend them without a primer. Telomerase, an enzyme with an inbuilt RNA template, extends the ends by copying the RNA template and extending one end of the chromosome. DNA polymerase can then extend the DNA using the primer. In this way, the ends of the chromosomes are protected. Cells have mechanisms for repairing DNA when it becomes damaged or errors are made in replication. These mechanisms include mismatch repair to replace nucleotides that are paired with a non-complementary base and nucleotide excision repair, which removes bases that are damaged such as thymine dimers.

Art Connections

Exercise:

Problem:

[link] You isolate a cell strain in which the joining together of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

Solution:

[link] Ligase, as this enzyme joins together Okazaki fragments.

Multiple Choice

Exercise:

Problem: DNA replicates by which of the following models?

a. conservative
b. semiconservative

c. dispersive
d. none of the above

Solution:

B

Exercise:

Problem:

The initial mechanism for repairing nucleotide errors in DNA is ________.

a. mismatch repair
b. DNA polymerase proofreading
c. nucleotide excision repair
d. thymine dimers

Solution:

B

Free Response

Exercise:

Problem:

How do the linear chromosomes in eukaryotes ensure that its ends are replicated completely?

Solution:

Telomerase has an inbuilt RNA template that extends the 3' end, so a primer is synthesized and extended. Thus, the ends are protected.

Glossary

DNA ligase
: the enzyme that catalyzes the joining of DNA fragments together

DNA polymerase
: an enzyme that synthesizes a new strand of DNA complementary to a template strand

helicase
: an enzyme that helps to open up the DNA helix during DNA replication by breaking the hydrogen bonds

lagging strand
: during replication of the 3' to 5' strand, the strand that is replicated in short fragments and away from the replication fork

leading strand
: the strand that is synthesized continuously in the 5' to 3' direction that is synthesized in the direction of the replication fork

mismatch repair
: a form of DNA repair in which non-complementary nucleotides are recognized, excised, and replaced with correct nucleotides

mutation
: a permanent variation in the nucleotide sequence of a genome

nucleotide excision repair
: a form of DNA repair in which the DNA molecule is unwound and separated in the region of the nucleotide damage, the damaged nucleotides are removed and replaced with new nucleotides using the complementary strand, and the DNA strand is resealed and allowed to rejoin its complement

Okazaki fragments
: the DNA fragments that are synthesized in short stretches on the lagging strand

primer
: a short stretch of RNA nucleotides that is required to initiate replication and allow DNA polymerase to bind and begin replication

replication fork
: the Y-shaped structure formed during the initiation of replication

semiconservative replication
: the method used to replicate DNA in which the double-stranded molecule is separated and each strand acts as a template for a new strand to be synthesized, so the resulting DNA molecules are composed of one new strand of nucleotides and one old strand of nucleotides

telomerase
: an enzyme that contains a catalytic part and an inbuilt RNA template; it functions to maintain telomeres at chromosome ends

telomere
: the DNA at the end of linear chromosomes

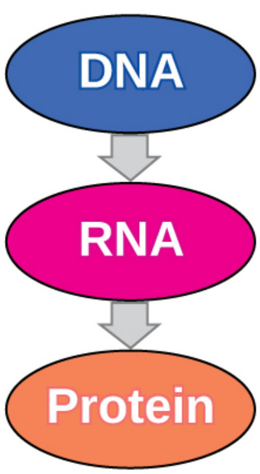
Transcription
By the end of this section, you will be able to:

- Explain the central dogma
- Explain the main steps of transcription
- Describe how eukaryotic mRNA is processed

In both prokaryotes and eukaryotes, the second function of DNA (the first was replication) is to provide the information needed to construct the proteins necessary so that the cell can perform all of its functions. To do this, the DNA is "read" or transcribed into an **mRNA** molecule. The mRNA then provides the code to form a protein by a process called translation. Through the processes of transcription and translation, a protein is built with a specific sequence of amino acids that was originally encoded in the DNA. This module discusses the details of transcription.

The Central Dogma: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the central dogma ([link]), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.



The central dogma states that

DNA encodes RNA, which in turn encodes protein.

The copying of DNA to mRNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every complementary nucleotide read in the DNA strand. The translation to protein is more complex because groups of three mRNA nucleotides correspond to one amino acid of the protein sequence. However, as we shall see in the next module, the translation to protein is still systematic, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.

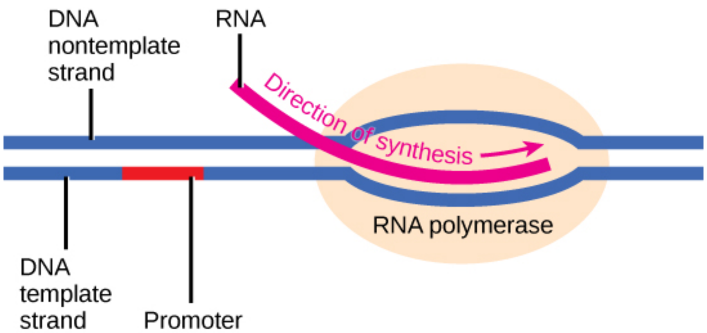
Transcription: from DNA to mRNA

Both prokaryotes and eukaryotes perform fundamentally the same process of transcription, with the important difference of the membrane-bound nucleus in eukaryotes. With the genes bound in the nucleus, transcription occurs in the nucleus of the cell and the mRNA transcript must be transported to the cytoplasm. The prokaryotes, which include bacteria and archaea, lack membrane-bound nuclei and other organelles, and transcription occurs in the cytoplasm of the cell. In both prokaryotes and eukaryotes, transcription occurs in three main stages: initiation, elongation, and termination.

Initiation

Transcription requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a **transcription bubble**. The DNA sequence onto which the proteins and enzymes involved in transcription bind to initiate the process is called a **promoter**. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it

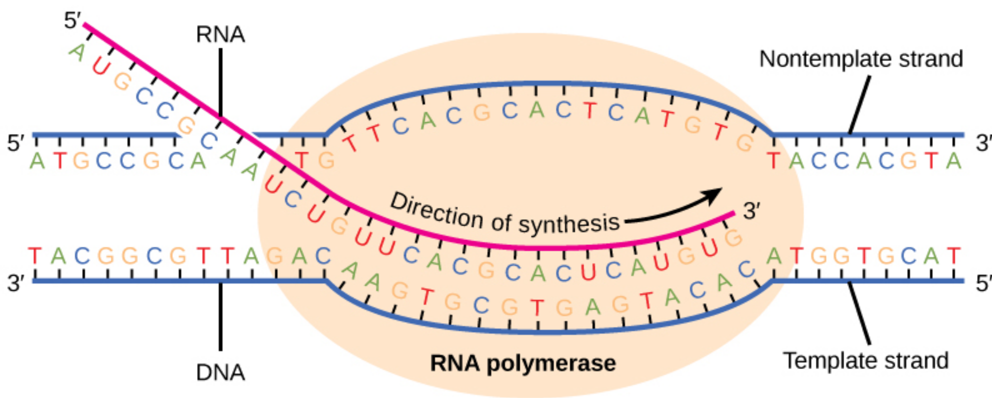
determines whether the corresponding gene is transcribed all of the time, some of the time, or hardly at all ([link]).



The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

Elongation

Transcription always proceeds from one of the two DNA strands, which is called the **template strand**. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **nontemplate strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. During elongation, an enzyme called **RNA polymerase** proceeds along the DNA template adding nucleotides by base pairing with the DNA template in a manner similar to DNA replication, with the difference that an RNA strand is being synthesized that does not remain bound to the DNA template. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it ([link]).

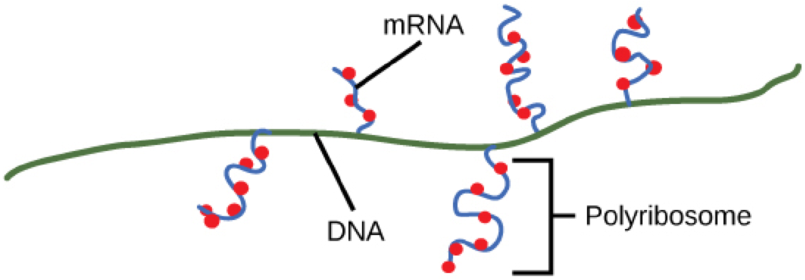


During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.

Termination

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals, but both involve repeated nucleotide sequences in the DNA template that result in RNA polymerase stalling, leaving the DNA template, and freeing the mRNA transcript.

On termination, the process of transcription is complete. In a prokaryotic cell, by the time termination occurs, the transcript would already have been used to partially synthesize numerous copies of the encoded protein because these processes can occur concurrently using multiple ribosomes (polyribosomes) ([link]). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.



Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

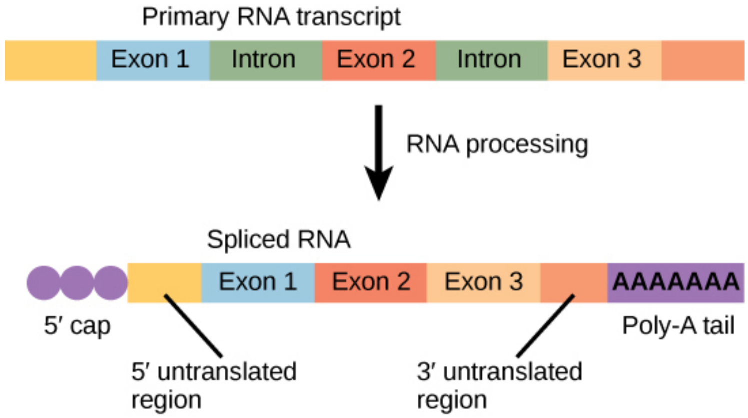
Eukaryotic RNA Processing

The newly transcribed eukaryotic mRNAs must undergo several processing steps before they can be transferred from the nucleus to the cytoplasm and translated into a protein. The additional steps involved in eukaryotic mRNA maturation create a molecule that is much more stable than a prokaryotic mRNA. For example, eukaryotic mRNAs last for several hours, whereas the typical prokaryotic mRNA lasts no more than five seconds.

The mRNA transcript is first coated in RNA-stabilizing proteins to prevent it from degrading while it is processed and exported out of the nucleus. This occurs while the pre-mRNA still is being synthesized by adding a special nucleotide “cap” to the 5' end of the growing transcript. In addition to preventing degradation, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

Once elongation is complete, an enzyme then adds a string of approximately 200 adenine residues to the 3' end, called the poly-A tail. This modification further protects the pre-mRNA from degradation and signals to cellular factors that the transcript needs to be exported to the cytoplasm.

Eukaryotic genes are composed of protein-coding sequences called **exons** (*ex*-on signifies that they are *ex*pressed) and *int*ervening sequences called **introns** (*int*-ron denotes their *int*ervening role). Introns are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins. It is essential that all of a pre-mRNA's introns be completely and precisely removed before protein synthesis so that the exons join together to code for the correct amino acids. If the process errs by even a single nucleotide, the sequence of the rejoined exons would be shifted, and the resulting protein would be nonfunctional. The process of removing introns and reconnecting exons is called **splicing** ([link]). Introns are removed and degraded while the pre-mRNA is still in the nucleus.



Eukaryotic mRNA contains introns that must be spliced out. A 5' cap and 3' tail are also added.

Section Summary

In prokaryotes, mRNA synthesis is initiated at a promoter sequence on the DNA template. Elongation synthesizes new mRNA. Termination liberates the mRNA and occurs by mechanisms that stall the RNA polymerase and cause it to fall off the DNA template. Newly transcribed eukaryotic mRNAs are modified with a cap and a poly-A tail. These structures protect the

mature mRNA from degradation and help export it from the nucleus. Eukaryotic mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs are exported from the nucleus to the cytoplasm.

Multiple Choice

Exercise:

Problem: A promoter is _________.

a. a specific sequence of DNA nucleotides
b. a specific sequence of RNA nucleotides
c. a protein that binds to DNA
d. an enzyme that synthesizes RNA

Solution:

A

Exercise:

Problem:

Portions of eukaryotic mRNA sequence that are removed during RNA processing are _________.

a. exons
b. caps
c. poly-A tails
d. introns

Solution:

D

Glossary

exon
: a sequence present in protein-coding mRNA after completion of pre-mRNA splicing

intron
: non–protein-coding intervening sequences that are spliced from mRNA during processing

mRNA
: messenger RNA; a form of RNA that carries the nucleotide sequence code for a protein sequence that is translated into a polypeptide sequence

nontemplate strand
: the strand of DNA that is not used to transcribe mRNA; this strand is identical to the mRNA except that T nucleotides in the DNA are replaced by U nucleotides in the mRNA

promoter
: a sequence on DNA to which RNA polymerase and associated factors bind and initiate transcription

RNA polymerase
: an enzyme that synthesizes an RNA strand from a DNA template strand

splicing
: the process of removing introns and reconnecting exons in a pre-mRNA

template strand
: the strand of DNA that specifies the complementary mRNA molecule

transcription bubble
: the region of locally unwound DNA that allows for transcription of mRNA

Translation

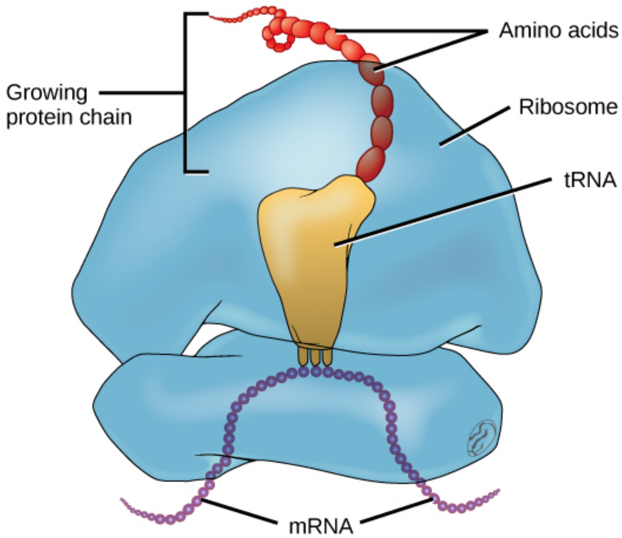
By the end of this section, you will be able to:

- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis
- Describe the genetic code and how the nucleotide sequence determines the amino acid and the protein sequence

The synthesis of proteins is one of a cell’s most energy-consuming metabolic processes. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform a wide variety of the functions of a cell. The process of translation, or protein synthesis, involves decoding an mRNA message into a polypeptide product. Amino acids are covalently strung together in lengths ranging from approximately 50 amino acids to more than 1,000.

The Protein Synthesis Machinery

In addition to the mRNA template, many other molecules contribute to the process of translation. The composition of each component may vary across species; for instance, ribosomes may consist of different numbers of ribosomal RNAs (**rRNA**) and polypeptides depending on the organism. However, the general structures and functions of the protein synthesis machinery are comparable from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymatic factors ([link]).



The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA. (credit: modification of work by NIGMS, NIH)

In *E. coli*, there are 200,000 ribosomes present in every cell at any given time. A ribosome is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. In eukaryotes, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

Ribosomes are located in the cytoplasm in prokaryotes and in the cytoplasm and endoplasmic reticulum of eukaryotes. Ribosomes are made up of a large and a small subunit that come together for translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds **tRNAs**, a type of RNA molecule that brings amino acids to the growing chain of the polypeptide. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction.

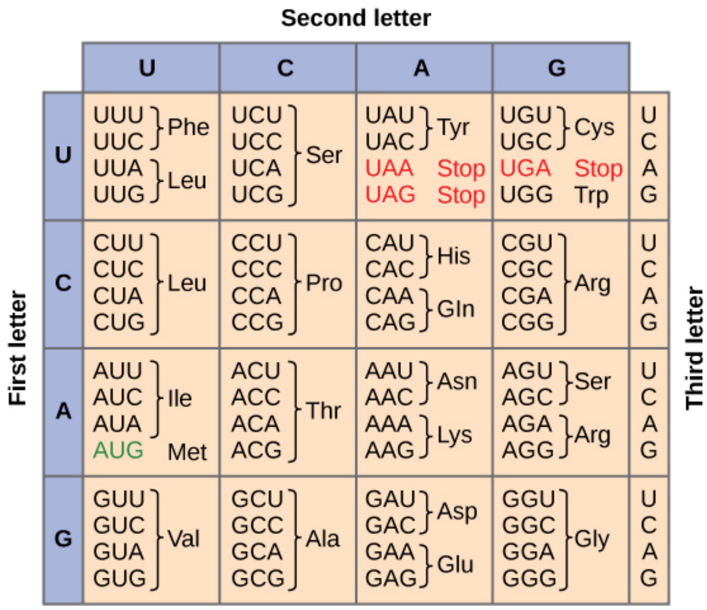
Depending on the species, 40 to 60 types of tRNA exist in the cytoplasm. Serving as adaptors, specific tRNAs bind to sequences on the mRNA

template and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually “translate” the language of RNA into the language of proteins. For each tRNA to function, it must have its specific amino acid bonded to it. In the process of tRNA “charging,” each tRNA molecule is bonded to its correct amino acid.

The Genetic Code

To summarize what we know to this point, the cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 letters. Each amino acid is defined by a three-nucleotide sequence called the triplet **codon**. The relationship between a nucleotide codon and its corresponding amino acid is called the **genetic code**.

Given the different numbers of “letters” in the mRNA and protein “alphabets,” combinations of nucleotides corresponded to single amino acids. Using a three-nucleotide code means that there are a total of 64 (4 × 4 × 4) possible combinations; therefore, a given amino acid is encoded by more than one nucleotide triplet ([link]).



This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

Three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **stop codons**. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the **start codon** to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. The genetic code is universal. With a few exceptions, virtually all species use the same genetic code for protein synthesis, which is powerful evidence that all life on Earth shares a common origin.

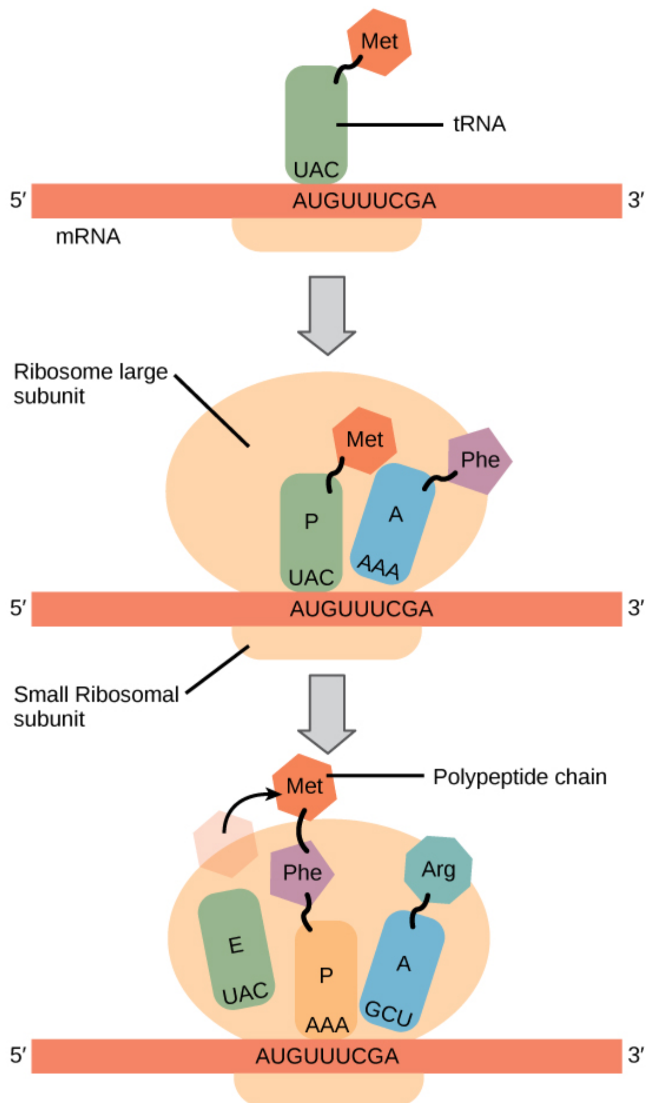
The Mechanism of Protein Synthesis

Just as with mRNA synthesis, protein synthesis can be divided into three phases: initiation, elongation, and termination. The process of translation is

similar in prokaryotes and eukaryotes. Here we will explore how translation occurs in *E. coli*, a representative prokaryote, and specify any differences between prokaryotic and eukaryotic translation.

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves the small ribosome subunit, the mRNA template, three initiation factors, and a special initiator tRNA. The initiator tRNA interacts with the AUG start codon, and links to a special form of the amino acid methionine that is typically removed from the polypeptide after translation is complete.

In prokaryotes and eukaryotes, the basics of polypeptide elongation are the same, so we will review elongation from the perspective of *E. coli*. The large ribosomal subunit of *E. coli* consists of three compartments: the A site binds incoming charged tRNAs (tRNAs with their attached specific amino acids). The P site binds charged tRNAs carrying amino acids that have formed bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E site releases dissociated tRNAs so they can be recharged with free amino acids. The ribosome shifts one codon at a time, catalyzing each process that occurs in the three sites. With each step, a charged tRNA enters the complex, the polypeptide becomes one amino acid longer, and an uncharged tRNA departs. The energy for each bond between amino acids is derived from GTP, a molecule similar to ATP ([link]). Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino acid polypeptide could be translated in just 10 seconds.



Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. When the ribosome encounters the stop codon, the growing polypeptide is released and the ribosome subunits dissociate and leave the mRNA. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

Note:
Concept in Action

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this site.

Section Summary

The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is the correspondence between the three-nucleotide mRNA codon and an amino acid. The genetic code is "translated" by the tRNA molecules, which associate a specific codon with a specific amino acid. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three stop codons. This means that more than one codon corresponds to an amino acid. Almost every species on the planet uses the same genetic code.

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template. Translation begins at the initiating AUG on the mRNA. The formation of bonds occurs between sequential amino acids specified by the mRNA template according to the genetic code. The ribosome accepts charged tRNAs, and as it steps along the mRNA, it catalyzes bonding between the new amino acid and the end of the growing polypeptide. The entire mRNA is translated in three-nucleotide “steps” of the ribosome. When a stop codon is encountered, a release factor binds and dissociates the components and frees the new protein.

Multiple Choice

Exercise:

Problem:

The RNA components of ribosomes are synthesized in the ________.

a. cytoplasm
b. nucleus
c. nucleolus
d. endoplasmic reticulum

Solution:

C

Exercise:

Problem:

How long would the peptide be that is translated from this MRNA sequence: 5'-AUGGGCUACCGA-3'?

a. 0
b. 2
c. 3

d. 4

Solution:

D

Free Response

Exercise:

Problem:

Transcribe and translate the following DNA sequence (nontemplate strand): 5'-ATGGCCGGTTATTAAGCA-3'

Solution:

The mRNA would be: 5'-AUGGCCGGUUAUUAAGCA-3'. The protein would be: MAGY. Even though there are six codons, the fifth codon corresponds to a stop, so the sixth codon would not be translated.

Glossary

codon
: three consecutive nucleotides in mRNA that specify the addition of a specific amino acid or the release of a polypeptide chain during translation

genetic code
: the amino acids that correspond to three-nucleotide codons of mRNA

rRNA
: ribosomal RNA; molecules of RNA that combine to form part of the ribosome

stop codon

one of the three mRNA codons that specifies termination of translation

start codon

the AUG (or, rarely GUG) on an mRNA from which translation begins; always specifies methionine

tRNA

transfer RNA; an RNA molecule that contains a specific three-nucleotide anticodon sequence to pair with the mRNA codon and also binds to a specific amino acid

How Genes Are Regulated

By the end of this section, you will be able to:

- Discuss why every cell does not express all of its genes
- Describe how prokaryotic gene expression occurs at the transcriptional level
- Understand that eukaryotic gene expression occurs at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

For a cell to function properly, necessary proteins must be synthesized at the proper time. All organisms and cells control or regulate the transcription and translation of their DNA into protein. The process of turning on a gene to produce RNA and protein is called **gene expression**. Whether in a simple unicellular organism or in a complex multicellular organism, each cell controls when and how its genes are expressed. For this to occur, there must be a mechanism to control when a gene is expressed to make RNA and protein, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

Cells in multicellular organisms are specialized; cells in different tissues look very different and perform different functions. For example, a muscle cell is very different from a liver cell, which is very different from a skin cell. These differences are a consequence of the expression of different sets of genes in each of these cells. All cells have certain basic functions they must perform for themselves, such as converting the energy in sugar molecules into energy in ATP. Each cell also has many genes that are not expressed, and expresses many that are not expressed by other cells, such that it can carry out its specialized functions. In addition, cells will turn on or off certain genes at different times in response to changes in the environment or at different times during the development of the organism. Unicellular organisms, both eukaryotic and prokaryotic, also turn on and off genes in response to the demands of their environment so that they can respond to special conditions.

The control of gene expression is extremely complex. Malfunctions in this process are detrimental to the cell and can lead to the development of many diseases, including cancer.

Prokaryotic versus Eukaryotic Gene Expression

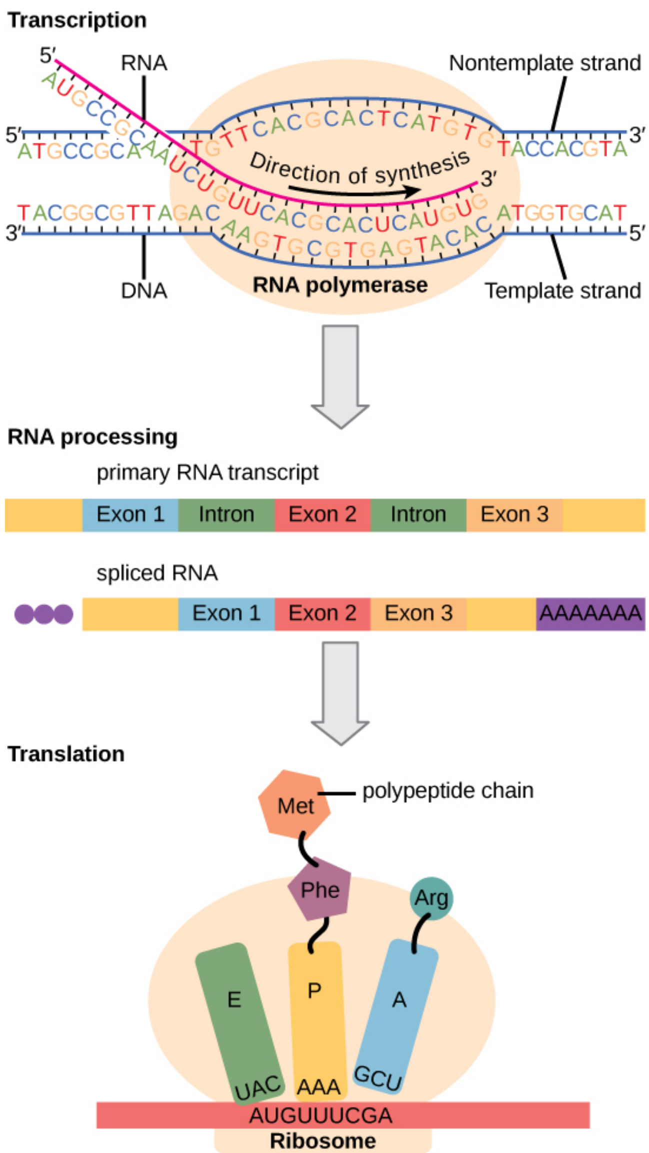
To understand how gene expression is regulated, we must first understand how a gene becomes a functional protein in a cell. The process occurs in both prokaryotic and eukaryotic cells, just in slightly different fashions.

Because prokaryotic organisms lack a cell nucleus, the processes of transcription and translation occur almost simultaneously. When the protein is no longer needed, transcription stops. As a result, the primary method to control what type and how much protein is expressed in a prokaryotic cell is through the regulation of DNA transcription into RNA. All the subsequent steps happen automatically. When more protein is required, more transcription occurs. Therefore, in prokaryotic cells, the control of gene expression is almost entirely at the transcriptional level.

The first example of such control was discovered using *E. coli* in the 1950s and 1960s by French researchers and is called the *lac* operon. The *lac* operon is a stretch of DNA with three adjacent genes that code for proteins that participate in the absorption and metabolism of lactose, a food source for *E. coli*. When lactose is not present in the bacterium's environment, the *lac* genes are transcribed in small amounts. When lactose is present, the genes are transcribed and the bacterium is able to use the lactose as a food source. The operon also contains a promoter sequence to which the RNA polymerase binds to begin transcription; between the promoter and the three genes is a region called the operator. When there is no lactose present, a protein known as a repressor binds to the operator and prevents RNA polymerase from binding to the promoter, except in rare cases. Thus very little of the protein products of the three genes is made. When lactose is present, an end product of lactose metabolism binds to the repressor protein and prevents it from binding to the operator. This allows RNA polymerase to bind to the promoter and freely transcribe the three genes, allowing the organism to metabolize the lactose.

Eukaryotic cells, in contrast, have intracellular organelles and are much more complex. Recall that in eukaryotic cells, the DNA is contained inside the cell's nucleus and it is transcribed into mRNA there. The newly synthesized mRNA is then transported out of the nucleus into the

cytoplasm, where ribosomes translate the mRNA into protein. The processes of transcription and translation are physically separated by the nuclear membrane; transcription occurs only within the nucleus, and translation only occurs outside the nucleus in the cytoplasm. The regulation of gene expression can occur at all stages of the process ([link]). Regulation may occur when the DNA is uncoiled and loosened from nucleosomes to bind transcription factors (**epigenetic** level), when the RNA is transcribed (transcriptional level), when RNA is processed and exported to the cytoplasm after it is transcribed (**post-transcriptional** level), when the RNA is translated into protein (translational level), or after the protein has been made (**post-translational** level).



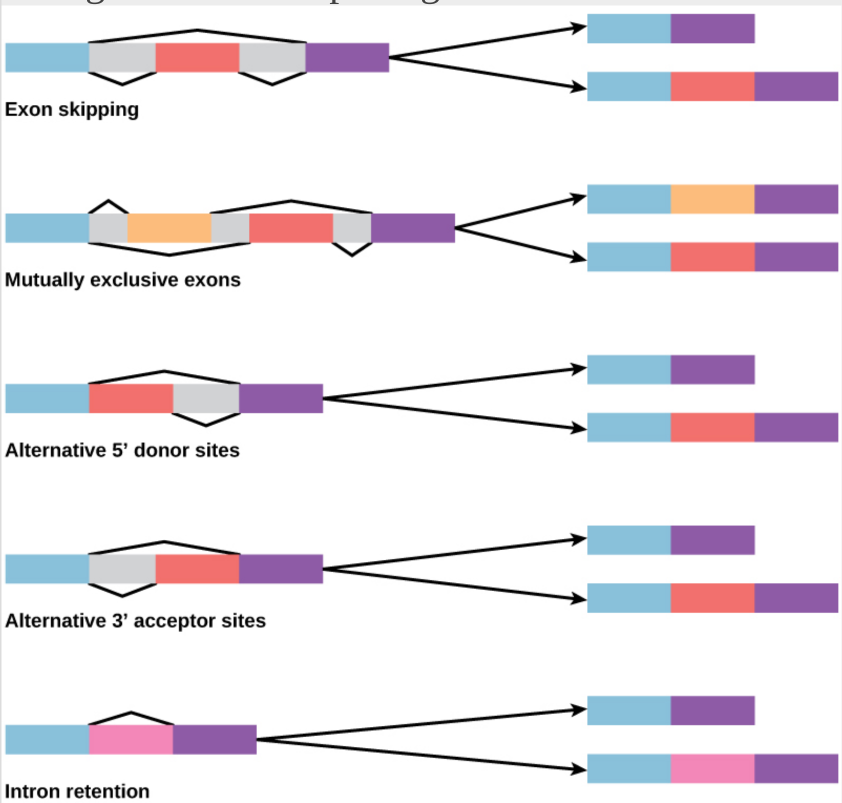
Eukaryotic gene expression is regulated during transcription and RNA processing, which take place in the nucleus, as well as during protein translation, which takes place in the cytoplasm. Further regulation may occur through post-translational modifications of proteins.

The differences in the regulation of gene expression between prokaryotes and eukaryotes are summarized in [link].

Differences in the Regulation of Gene Expression of Prokaryotic and Eukaryotic Organisms	
Prokaryotic organisms	**Eukaryotic organisms**
Lack nucleus	Contain nucleus
RNA transcription and protein translation occur almost simultaneously	• RNA transcription occurs prior to protein translation, and it takes place in the nucleus. RNA translation to protein occurs in the cytoplasm. • RNA post-processing includes addition of a 5' cap, poly-A tail, and excision of introns and splicing of exons.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels (epigenetic, transcriptional, post-transcriptional, translational, and post-translational)

Note:
Evolution in Action
Alternative RNA Splicing

In the 1970s, genes were first observed that exhibited **alternative RNA splicing**. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different combinations of introns (and sometimes exons) are removed from the transcript ([link]). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of gene regulation, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells, or at different stages of development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70% of genes in humans are expressed as multiple proteins through alternative splicing.



There are five basic modes of alternative splicing. Segments of pre-mRNA with exons shown in blue, red, orange, and pink can be spliced to produce a variety of new mature mRNA segments.

How could alternative splicing evolve? Introns have a beginning and ending recognition sequence, and it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such exon skipping, but mutations are likely to lead to their failure. Such “mistakes” would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is alternative splicing rather than mutations in a sequence. However, alternative splicing would create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way—by providing genes that may evolve without eliminating the original functional protein.

Section Summary

While all somatic cells within an organism contain the same DNA, not all cells within that organism express the same proteins. Prokaryotic organisms express the entire DNA they encode in every cell, but not necessarily all at the same time. Proteins are expressed only when they are needed. Eukaryotic organisms express a subset of the DNA that is encoded in any given cell. In each cell type, the type and amount of protein is regulated by controlling gene expression. To express a protein, the DNA is first transcribed into RNA, which is then translated into proteins. In prokaryotic cells, these processes occur almost simultaneously. In eukaryotic cells, transcription occurs in the nucleus and is separate from the translation that occurs in the cytoplasm. Gene expression in prokaryotes is regulated only at the transcriptional level, whereas in eukaryotic cells, gene expression is regulated at the epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels.

Multiple Choice

Exercise:

Problem:

Control of gene expression in eukaryotic cells occurs at which level(s)?

a. only the transcriptional level
b. epigenetic and transcriptional levels
c. epigenetic, transcriptional, and translational levels
d. epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels

---

Solution:

D

Exercise:

Problem: Post-translational control refers to:

a. regulation of gene expression after transcription
b. regulation of gene expression after translation
c. control of epigenetic activation
d. period between transcription and translation

Solution:

B

Free Response

Exercise:

Problem:

Describe how controlling gene expression will alter the overall protein levels in the cell.

Solution:

The cell controls which protein is expressed, and to what level that protein is expressed, in the cell. Prokaryotic cells alter the transcription rate to turn genes on or off. This method will increase or decrease protein levels in response to what is needed by the cell. Eukaryotic cells change the accessibility (epigenetic), transcription, or translation of a gene. This will alter the amount of RNA, and the lifespan of the RNA, to alter the amount of protein that exists. Eukaryotic cells also change the protein’s translation to increase or decrease its overall levels. Eukaryotic organisms are much more complex and can manipulate protein levels by changing many stages in the process.

Glossary

alternative RNA splicing
: a post-transcriptional gene regulation mechanism in eukaryotes in which multiple protein products are produced by a single gene through alternative splicing combinations of the RNA transcript

epigenetic
: describing non-genetic regulatory factors, such as changes in modifications to histone proteins and DNA that control accessibility to genes in chromosomes

gene expression
: processes that control whether a gene is expressed

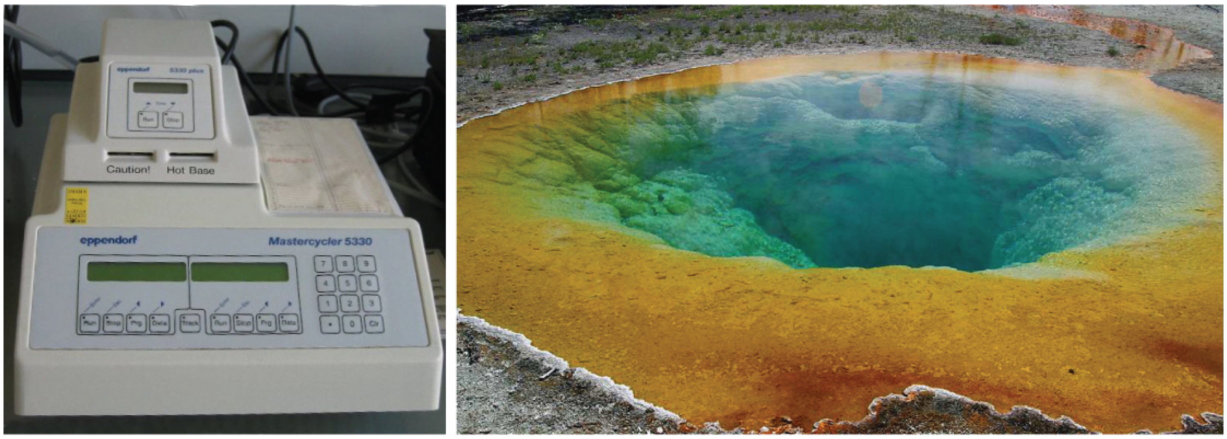
post-transcriptional
: control of gene expression after the RNA molecule has been created but before it is translated into protein

post-translational
: control of gene expression after a protein has been created

Introduction class="introduction"

(a) A thermal cycler, such as the one shown here, is a basic tool used to study DNA in a process called the polymerase chain reaction (PCR). The polymerase enzyme most often used with PCR comes from a strain of bacteria that lives in (b) the hot springs of Yellowstone National Park. (credit a: modification of work by Magnus Manske; credit b:

modification of work by Jon Sullivan)



(a) (b)

The latter half of the twentieth century began with the discovery of the structure of DNA, then progressed to the development of the basic tools used to study and manipulate DNA. These advances, as well as advances in our understanding of and ability to manipulate cells, have led some to refer to the twenty-first century as the biotechnology century. The rate of discovery and of the development of new applications in medicine, agriculture, and energy is expected to accelerate, bringing huge benefits to humankind and perhaps also significant risks. Many of these developments are expected to raise significant ethical and social questions that human societies have not yet had to consider.

Cloning and Genetic Engineering
By the end of this section, you will be able to:

- Explain the basic techniques used to manipulate genetic material
- Explain molecular and reproductive cloning

Biotechnology is the use of artificial methods to modify the genetic material of living organisms or cells to produce novel compounds or to perform new functions. Biotechnology has been used for improving livestock and crops since the beginning of agriculture through selective breeding. Since the discovery of the structure of DNA in 1953, and particularly since the development of tools and methods to manipulate DNA in the 1970s, biotechnology has become synonymous with the manipulation of organisms' DNA at the molecular level. The primary applications of this technology are in medicine (for the production of vaccines and antibiotics) and in agriculture (for the genetic modification of crops). Biotechnology also has many industrial applications, such as fermentation, the treatment of oil spills, and the production of biofuels, as well as many household applications such as the use of enzymes in laundry detergent.

Manipulating Genetic Material

To accomplish the applications described above, biotechnologists must be able to extract, manipulate, and analyze nucleic acids.

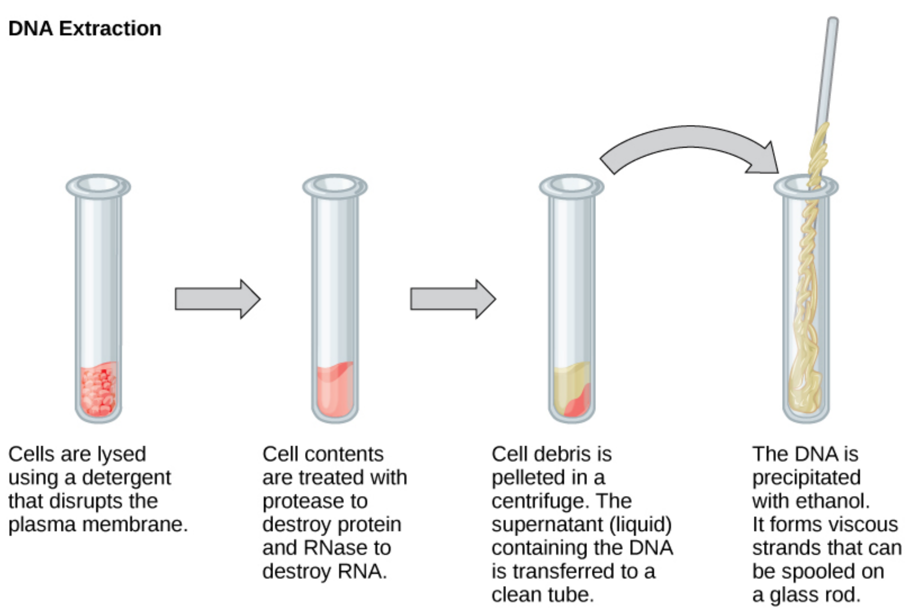
Review of Nucleic Acid Structure

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a nitrogenous base). The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus of eukaryotic organisms is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases.

Unlike DNA in eukaryotic cells, RNA molecules leave the nucleus. Messenger RNA (mRNA) is analyzed most frequently because it represents the protein-coding genes that are being expressed in the cell.

Isolation of Nucleic Acids

To study or manipulate nucleic acids, the DNA must first be extracted from cells. Various techniques are used to extract different types of DNA ([link]). Most nucleic acid extraction techniques involve steps to break open the cell, and then the use of enzymatic reactions to destroy all undesired macromolecules. Cells are broken open using a detergent solution containing buffering compounds. To prevent degradation and contamination, macromolecules such as proteins and RNA are inactivated using enzymes. The DNA is then brought out of solution using alcohol. The resulting DNA, because it is made up of long polymers, forms a gelatinous mass.

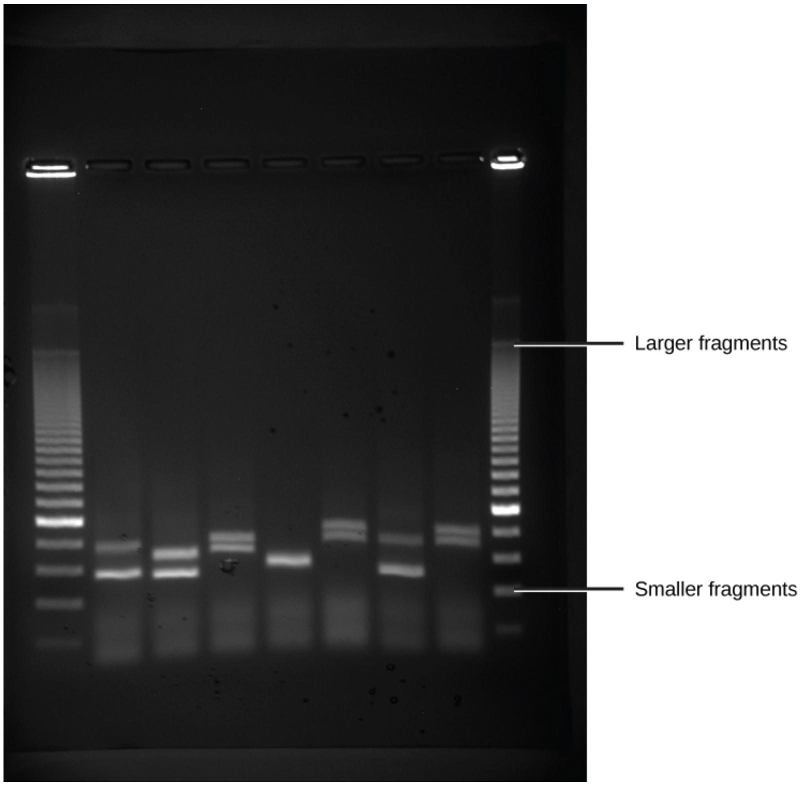


This diagram shows the basic method used for the extraction of DNA.

RNA is studied to understand gene expression patterns in cells. RNA is naturally very unstable because enzymes that break down RNA are commonly present in nature. Some are even secreted by our own skin and are very difficult to inactivate. Similar to DNA extraction, RNA extraction involves the use of various buffers and enzymes to inactivate other macromolecules and preserve only the RNA.

Gel Electrophoresis

Because nucleic acids are negatively charged ions at neutral or alkaline pH in an aqueous environment, they can be moved by an electric field. **Gel electrophoresis** is a technique used to separate charged molecules on the basis of size and charge. The nucleic acids can be separated as whole chromosomes or as fragments. The nucleic acids are loaded into a slot at one end of a gel matrix, an electric current is applied, and negatively charged molecules are pulled toward the opposite end of the gel (the end with the positive electrode). Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. The nucleic acids in a gel matrix are invisible until they are stained with a compound that allows them to be seen, such as a dye. Distinct fragments of nucleic acids appear as bands at specific distances from the top of the gel (the negative electrode end) that are based on their size ([link]). A mixture of many fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.

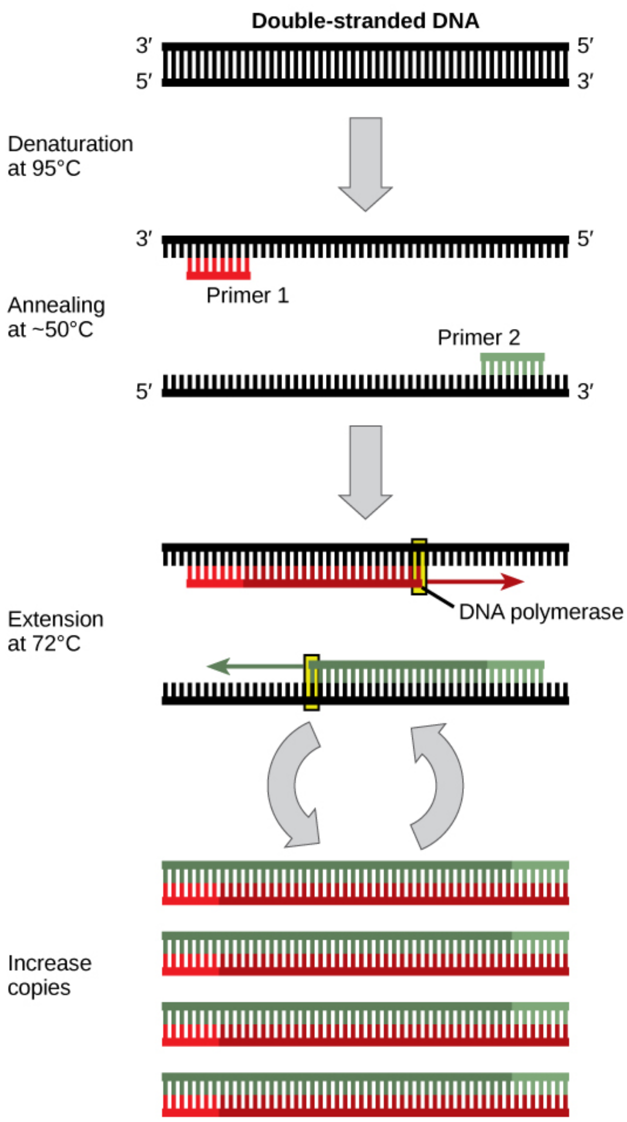


Shown are DNA fragments from six samples run on a gel, stained with a fluorescent dye and viewed under UV light. (credit: modification of work by James Jacob, Tompkins Cortland Community College)

Polymerase Chain Reaction

DNA analysis often requires focusing on one or more specific regions of the genome. It also frequently involves situations in which only one or a few copies of a DNA molecule are available for further analysis. These amounts are insufficient for most procedures, such as gel electrophoresis. **Polymerase chain reaction (PCR)** is a technique used to rapidly increase the number of copies of specific regions of DNA for further analyses ([link]). PCR uses a special form of DNA polymerase, the enzyme that

replicates DNA, and other short nucleotide sequences called primers that base pair to a specific portion of the DNA being replicated. PCR is used for many purposes in laboratories. These include: 1) the identification of the owner of a DNA sample left at a crime scene; 2) paternity analysis; 3) the comparison of small amounts of ancient DNA with modern organisms; and 4) determining the sequence of nucleotides in a specific region.



Polymerase chain reaction, or PCR, is used to produce many copies of a specific sequence of

DNA using a special form of DNA polymerase.

Cloning

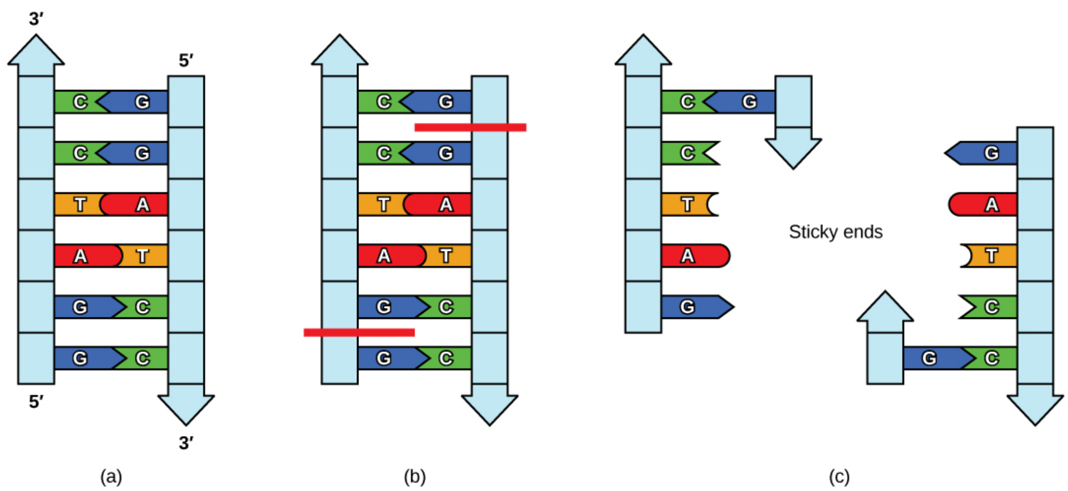
In general, **cloning** means the creation of a perfect replica. Typically, the word is used to describe the creation of a genetically identical copy. In biology, the re-creation of a whole organism is referred to as “reproductive cloning.” Long before attempts were made to clone an entire organism, researchers learned how to copy short stretches of DNA—a process that is referred to as molecular cloning.

Molecular Cloning

Cloning allows for the creation of multiple copies of genes, expression of genes, and study of specific genes. To get the DNA fragment into a bacterial cell in a form that will be copied or expressed, the fragment is first inserted into a plasmid. A **plasmid** (also called a vector in this context) is a small circular DNA molecule that replicates independently of the chromosomal DNA in bacteria. In cloning, the plasmid molecules can be used to provide a "vehicle" in which to insert a desired DNA fragment. Modified plasmids are usually reintroduced into a bacterial host for replication. As the bacteria divide, they copy their own DNA (including the plasmids). The inserted DNA fragment is copied along with the rest of the bacterial DNA. In a bacterial cell, the fragment of DNA from the human genome (or another organism that is being studied) is referred to as foreign DNA to differentiate it from the DNA of the bacterium (the host DNA).

Plasmids occur naturally in bacterial populations (such as *Escherichia coli*) and have genes that can contribute favorable traits to the organism, such as antibiotic resistance (the ability to be unaffected by antibiotics). Plasmids have been highly engineered as vectors for molecular cloning and for the subsequent large-scale production of important molecules, such as insulin.

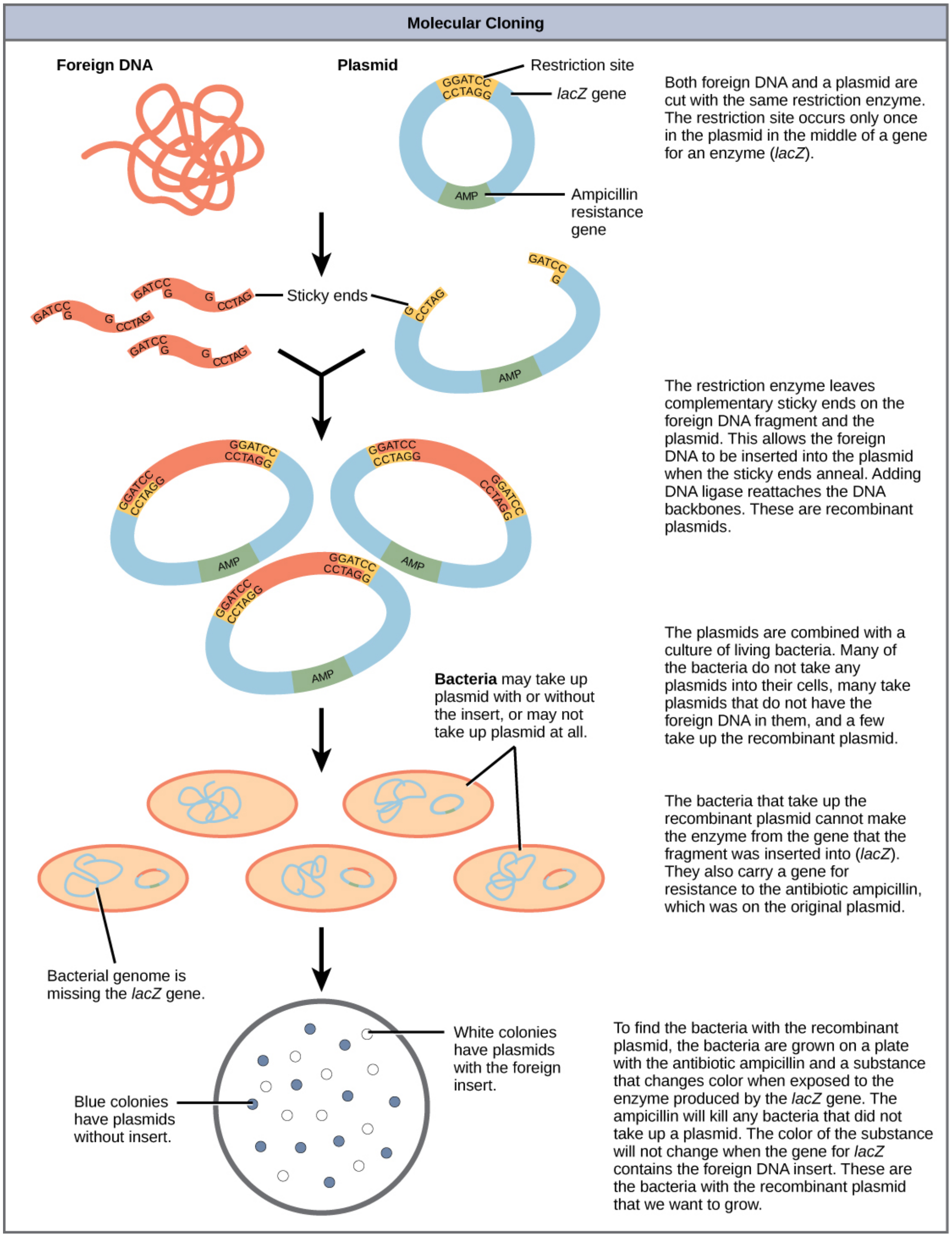
A valuable characteristic of plasmid vectors is the ease with which a foreign DNA fragment can be introduced. These plasmid vectors contain many short DNA sequences that can be cut with different commonly available **restriction enzymes**. Restriction enzymes (also called restriction endonucleases) recognize specific DNA sequences and cut them in a predictable manner; they are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction enzymes make staggered cuts in the two strands of DNA, such that the cut ends have a 2- to 4-nucleotide single-stranded overhang. The sequence that is recognized by the restriction enzyme is a four- to eight-nucleotide sequence that is a palindrome. Like with a word palindrome, this means the sequence reads the same forward and backward. In most cases, the sequence reads the same forward on one strand and backward on the complementary strand. When a staggered cut is made in a sequence like this, the overhangs are complementary ([link]).



In this (a) six-nucleotide restriction enzyme recognition site, notice that the sequence of six nucleotides reads the same in the 5' to 3' direction on one strand as it does in the 5' to 3' direction on the complementary strand. This is known as a palindrome. (b) The restriction enzyme makes breaks in the DNA strands, and (c) the cut in the DNA results in “sticky ends”. Another piece of DNA cut on either end by the same

restriction enzyme could attach to these sticky ends and be inserted into the gap made by this cut.

Because these overhangs are capable of coming back together by hydrogen bonding with complementary overhangs on a piece of DNA cut with the same restriction enzyme, these are called “sticky ends.” The process of forming hydrogen bonds between complementary sequences on single strands to form double-stranded DNA is called **annealing**. Addition of an enzyme called DNA ligase, which takes part in DNA replication in cells, permanently joins the DNA fragments when the sticky ends come together. In this way, any DNA fragment can be spliced between the two ends of a plasmid DNA that has been cut with the same restriction enzyme ([link]).



This diagram shows the steps involved in molecular cloning.

Plasmids with foreign DNA inserted into them are called **recombinant DNA** molecules because they contain new combinations of genetic material. Proteins that are produced from recombinant DNA molecules are called **recombinant proteins**. Not all recombinant plasmids are capable of expressing genes. Plasmids may also be engineered to express proteins only when stimulated by certain environmental factors, so that scientists can control the expression of the recombinant proteins.

Reproductive Cloning

Reproductive cloning is a method used to make a clone or an identical copy of an entire multicellular organism. Most multicellular organisms undergo reproduction by sexual means, which involves the contribution of DNA from two individuals (parents), making it impossible to generate an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to reproductively clone mammals in the laboratory.

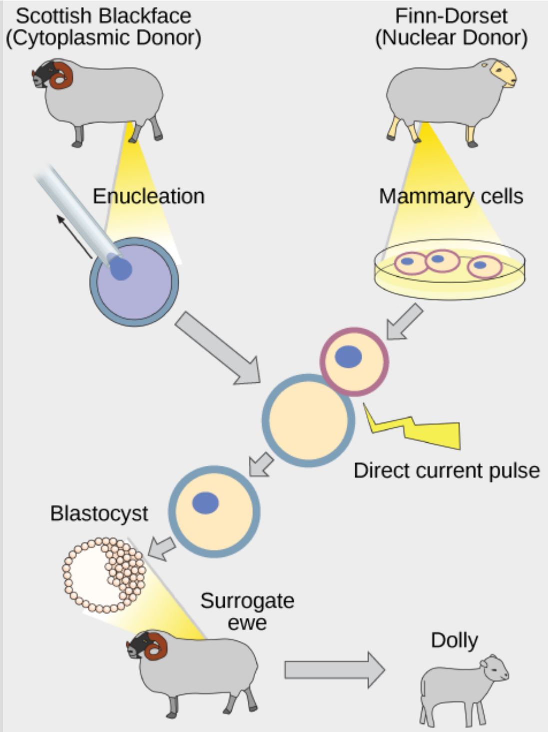
Natural sexual reproduction involves the union, during fertilization, of a sperm and an egg. Each of these gametes is haploid, meaning they contain one set of chromosomes in their nuclei. The resulting cell, or zygote, is then diploid and contains two sets of chromosomes. This cell divides mitotically to produce a multicellular organism. However, the union of just any two cells cannot produce a viable zygote; there are components in the cytoplasm of the egg cell that are essential for the early development of the embryo during its first few cell divisions. Without these provisions, there would be no subsequent development. Therefore, to produce a new individual, both a diploid genetic complement and an egg cytoplasm are required. The approach to producing an artificially cloned individual is to take the egg cell of one individual and to remove the haploid nucleus. Then a diploid nucleus from a body cell of a second individual, the donor, is put into the egg cell. The egg is then stimulated to divide so that development proceeds. This sounds simple, but in fact it takes many attempts before each of the steps is completed successfully.

The first cloned agricultural animal was Dolly, a sheep who was born in 1996. The success rate of reproductive cloning at the time was very low.

Dolly lived for six years and died of a lung tumor ([link]). There was speculation that because the cell DNA that gave rise to Dolly came from an older individual, the age of the DNA may have affected her life expectancy. Since Dolly, several species of animals (such as horses, bulls, and goats) have been successfully cloned.

There have been attempts at producing cloned human embryos as sources of embryonic stem cells. In the procedure, the DNA from an adult human is introduced into a human egg cell, which is then stimulated to divide. The technology is similar to the technology that was used to produce Dolly, but the embryo is never implanted into a surrogate mother. The cells produced are called embryonic stem cells because they have the capacity to develop into many different kinds of cells, such as muscle or nerve cells. The stem cells could be used to research and ultimately provide therapeutic applications, such as replacing damaged tissues. The benefit of cloning in this instance is that the cells used to regenerate new tissues would be a perfect match to the donor of the original DNA. For example, a leukemia patient would not require a sibling with a tissue match for a bone-marrow transplant.

Note:
Art Connection



Dolly the sheep was the first agricultural animal to be cloned. To create Dolly, the nucleus was removed from a donor egg cell. The enucleated egg was placed next to the other cell, then they were shocked to fuse. They were shocked again to start division. The cells were allowed to divide for several days until an early embryonic stage was reached, before being implanted in a surrogate mother.

Why was Dolly a Finn-Dorset and not a Scottish Blackface sheep?

Genetic Engineering

Using recombinant DNA technology to modify an organism’s DNA to achieve desirable traits is called **genetic engineering**. Addition of foreign DNA in the form of recombinant DNA vectors that are generated by molecular cloning is the most common method of genetic engineering. An organism that receives the recombinant DNA is called a **genetically modified organism** (GMO). If the foreign DNA that is introduced comes from a different species, the host organism is called **transgenic**. Bacteria, plants, and animals have been genetically modified since the early 1970s for academic, medical, agricultural, and industrial purposes. These applications will be examined in more detail in the next module.

Note:
Concept in Action

Watch this short video explaining how scientists create a transgenic animal.

Although the classic methods of studying the function of genes began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: "What does this gene or DNA element do?" This technique, called **reverse genetics**, has resulted in reversing the classical genetic methodology. One example of this method is analogous to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the wing’s function is flight. The classic genetic method compares insects that cannot fly with insects that can fly, and observes that the non-flying insects have lost wings. Similarly in a reverse genetics

approach, mutating or deleting genes provides researchers with clues about gene function. Alternately, reverse genetics can be used to cause a gene to overexpress itself to determine what phenotypic effects may occur.

Section Summary

Nucleic acids can be isolated from cells for the purposes of further analysis by breaking open the cells and enzymatically destroying all other major macromolecules. Fragmented or whole chromosomes can be separated on the basis of size by gel electrophoresis. Short stretches of DNA can be amplified by PCR. DNA can be cut (and subsequently re-spliced together) using restriction enzymes. The molecular and cellular techniques of biotechnology allow researchers to genetically engineer organisms, modifying them to achieve desirable traits.

Cloning may involve cloning small DNA fragments (molecular cloning), or cloning entire organisms (reproductive cloning). In molecular cloning with bacteria, a desired DNA fragment is inserted into a bacterial plasmid using restriction enzymes and the plasmid is taken up by a bacterium, which will then express the foreign DNA. Using other techniques, foreign genes can be inserted into eukaryotic organisms. In each case, the organisms are called transgenic organisms. In reproductive cloning, a donor nucleus is put into an enucleated egg cell, which is then stimulated to divide and develop into an organism.

In reverse genetics methods, a gene is mutated or removed in some way to identify its effect on the phenotype of the whole organism as a way to determine its function.

Art Connections

Exercise:

Problem:

[link] Why was Dolly a Finn-Dorset and not a Scottish Blackface sheep?

Solution:

[link] Because even though the original cell came from a Scottish Blackface sheep and the surrogate mother was a Scottish Blackface, the DNA came from a Finn-Dorset.

Multiple Choice

Exercise:

Problem:

In gel electrophoresis of DNA, the different bands in the final gel form because the DNA molecules _________.

a. are from different organisms
b. have different lengths
c. have different nucleotide compositions
d. have different genes

Solution:

B

Exercise:

Problem:

In the reproductive cloning of an animal, the genome of the cloned individual comes from _________.

a. a sperm cell
b. an egg cell
c. any gamete cell
d. a body cell

Solution:

D

Exercise:

Problem: What carries a gene from one organism into a bacteria cell?

a. a plasmid
b. an electrophoresis gel
c. a restriction enzyme
d. polymerase chain reaction

Solution:

A

Free Response

Exercise:

Problem:

What is the purpose and benefit of the polymerase chain reaction?

Solution:

The polymerase chain reaction is used to quickly produce many copies of a specific segment of DNA when only one or a very few copies are originally present. The benefit of PCR is that there are many instances in which we would like to know something about a sample of DNA when only very small amounts are available. PCR allows us to increase the number of DNA molecules so that other tests, such as sequencing, can be performed with it.

Glossary

anneal

in molecular biology, the process by which two single strands of DNA hydrogen bond at complementary nucleotides to form a double-stranded molecule

biotechnology
: the use of artificial methods to modify the genetic material of living organisms or cells to produce novel compounds or to perform new functions

cloning
: the production of an exact copy—specifically, an exact genetic copy—of a gene, cell, or organism

gel electrophoresis
: a technique used to separate molecules on the basis of their ability to migrate through a semisolid gel in response to an electric current

genetic engineering
: alteration of the genetic makeup of an organism using the molecular methods of biotechnology

genetically modified organism (GMO)
: an organism whose genome has been artificially changed

plasmid
: a small circular molecule of DNA found in bacteria that replicates independently of the main bacterial chromosome; plasmids code for some important traits for bacteria and can be used as vectors to transport DNA into bacteria in genetic engineering applications

polymerase chain reaction (PCR)
: a technique used to make multiple copies of DNA

recombinant DNA
: a combination of DNA fragments generated by molecular cloning that does not exist in nature

recombinant protein

a protein that is expressed from recombinant DNA molecules

restriction enzyme
: an enzyme that recognizes a specific nucleotide sequence in DNA and cuts the DNA double strand at that recognition site, often with a staggered cut leaving short single strands or “sticky” ends

reverse genetics
: a form of genetic analysis that manipulates DNA to disrupt or affect the product of a gene to analyze the gene’s function

reproductive cloning
: cloning of entire organisms

transgenic
: describing an organism that receives DNA from a different species

Biotechnology in Medicine and Agriculture
By the end of this section, you will be able to:

- Describe uses of biotechnology in medicine
- Describe uses of biotechnology in agriculture

It is easy to see how biotechnology can be used for medicinal purposes. Knowledge of the genetic makeup of our species, the genetic basis of heritable diseases, and the invention of technology to manipulate and fix mutant genes provides methods to treat diseases. Biotechnology in agriculture can enhance resistance to disease, pests, and environmental stress to improve both crop yield and quality.

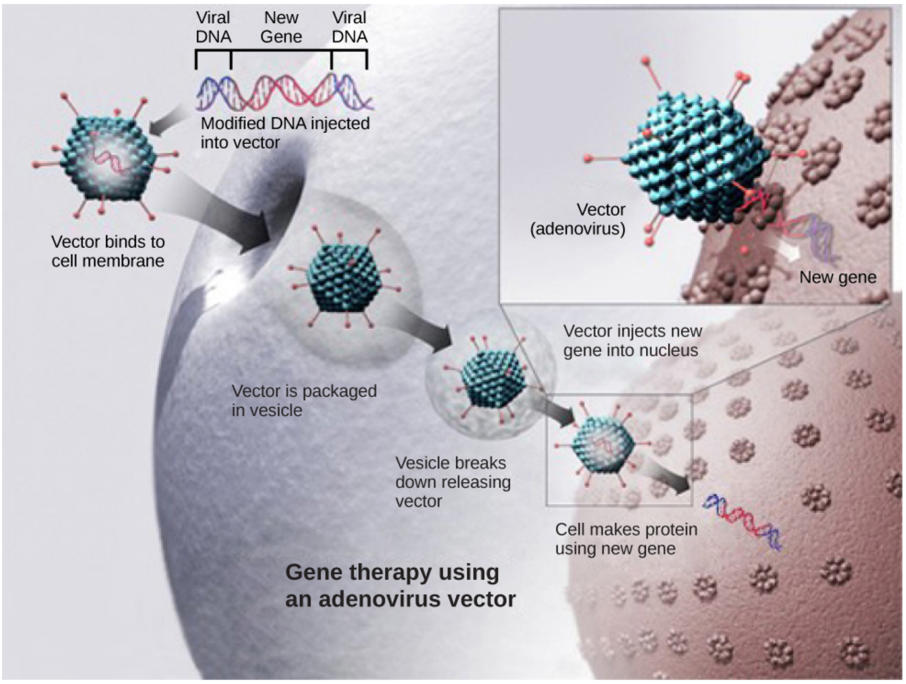
Genetic Diagnosis and Gene Therapy

The process of testing for suspected genetic defects before administering treatment is called genetic diagnosis by genetic testing. In some cases in which a genetic disease is present in an individual's family, family members may be advised to undergo genetic testing. For example, mutations in the *BRCA* genes may increase the likelihood of developing breast and ovarian cancers in women and some other cancers in women and men. A woman with breast cancer can be screened for these mutations. If one of the high-risk mutations is found, her female relatives may also wish to be screened for that particular mutation, or simply be more vigilant for the occurrence of cancers. Genetic testing is also offered for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

Note:
Concept in Action

See how human DNA is extracted for uses such as genetic testing.

Gene therapy is a genetic engineering technique that may one day be used to cure certain genetic diseases. In its simplest form, it involves the introduction of a non-mutated gene at a random location in the genome to cure a disease by replacing a protein that may be absent in these individuals because of a genetic mutation. The non-mutated gene is usually introduced into diseased cells as part of a vector transmitted by a virus, such as an adenovirus, that can infect the host cell and deliver the foreign DNA into the genome of the targeted cell ([link]). To date, gene therapies have been primarily experimental procedures in humans. A few of these experimental treatments have been successful, but the methods may be important in the future as the factors limiting its success are resolved.



This diagram shows the steps involved in curing disease with gene therapy using an adenovirus vector. (credit: modification of work by NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms or viruses to stimulate the immune system. Modern techniques use specific genes of microorganisms cloned into vectors and mass-produced in bacteria to make large quantities of specific substances to stimulate the immune system. The substance is then used as a vaccine. In some cases, such as the H1N1 flu vaccine, genes cloned from the virus have been used to combat the constantly changing strains of this virus.

Antibiotics kill bacteria and are naturally produced by microorganisms such as fungi; penicillin is perhaps the most well-known example. Antibiotics are produced on a large scale by cultivating and manipulating fungal cells. The fungal cells have typically been genetically modified to improve the yields of the antibiotic compound.

Recombinant DNA technology was used to produce large-scale quantities of the human hormone insulin in *E. coli* as early as 1978. Previously, it was only possible to treat diabetes with pig insulin, which caused allergic reactions in many humans because of differences in the insulin molecule. In addition, human growth hormone (HGH) is used to treat growth disorders in children. The HGH gene was cloned from a cDNA (complementary DNA) library and inserted into *E. coli* cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins used in medicine are successfully produced in bacteria, some proteins need a eukaryotic animal host for proper processing. For this reason, genes have been cloned and expressed in animals such as sheep, goats, chickens, and mice. Animals that have been modified to express recombinant DNA are called transgenic animals ([link]).

It can be seen that two of these mice are transgenic because they have a gene that causes them to fluoresce under a UV light. The non-transgenic mouse does not have the gene that

causes fluorescence. (credit: Ingrid Moen et al.)

Several human proteins are expressed in the milk of transgenic sheep and goats. In one commercial example, the FDA has approved a blood anticoagulant protein that is produced in the milk of transgenic goats for use in humans. Mice have been used extensively for expressing and studying the effects of recombinant genes and mutations.

Transgenic Plants

Manipulating the DNA of plants (creating genetically modified organisms, or GMOs) has helped to create desirable traits such as disease resistance, herbicide, and pest resistance, better nutritional value, and better shelf life ([link]). Plants are the most important source of food for the human population. Farmers developed ways to select for plant varieties with desirable traits long before modern-day biotechnology practices were established.

Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

Transgenic plants have received DNA from other species. Because they contain unique combinations of genes and are not restricted to the laboratory, transgenic plants and other GMOs are closely monitored by government agencies to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because foreign genes can spread to other species in the environment, particularly in the pollen and seeds of plants, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants to be genetically engineered.

Transformation of Plants Using *Agrobacterium tumefaciens*

In plants, tumors caused by the bacterium *Agrobacterium tumefaciens* occur by transfer of DNA from the bacterium to the plant. The artificial introduction of DNA into plant cells is more challenging than in animal cells because of the thick plant cell wall. Researchers used the natural transfer of DNA from *Agrobacterium* to a plant host to introduce DNA fragments of their choice into plant hosts. In nature, the disease-causing *A. tumefaciens* have a set of plasmids that contain genes that integrate into the infected plant cell's genome. Researchers manipulate the plasmids to carry the desired DNA fragment and insert it into the plant genome.

The Organic Insecticide *Bacillus thuringiensis*

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals that are toxic to many insect species that feed on plants. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin is activated in the intestines of the insects, death occurs within a couple of days. The crystal toxin genes have been cloned from the bacterium and introduced into plants, therefore allowing plants to produce their own crystal Bt toxin that acts against insects. Bt toxin is safe for the environment and non-toxic to mammals (including humans). As a result, it has been approved for use by organic farmers as a natural insecticide. There is some concern, however, that insects may evolve resistance to the Bt toxin in the same way that bacteria evolve resistance to antibiotics.

FlavrSavr Tomato

The first GM crop to be introduced into the market was the FlavrSavr Tomato produced in 1994. Molecular genetic technology was used to slow down the process of softening and rotting caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the flavor of this tomato. The FlavrSavr tomato did

not successfully stay in the market because of problems maintaining and shipping the crop.

Section Summary

Genetic testing is performed to identify disease-causing genes, and can be used to benefit affected individuals and their relatives who have not developed disease symptoms yet. Gene therapy—by which functioning genes are incorporated into the genomes of individuals with a non-functioning mutant gene—has the potential to cure heritable diseases. Transgenic organisms possess DNA from a different species, usually generated by molecular cloning techniques. Vaccines, antibiotics, and hormones are examples of products obtained by recombinant DNA technology. Transgenic animals have been created for experimental purposes and some are used to produce some human proteins.

Genes are inserted into plants, using plasmids in the bacterium *Agrobacterium tumefaciens*, which infects plants. Transgenic plants have been created to improve the characteristics of crop plants—for example, by giving them insect resistance by inserting a gene for a bacterial toxin.

Multiple Choice

Exercise:

Problem: What is a genetically modified organism (GMO)?

a. a plant with certain genes removed
b. an organism with an artificially altered genome
c. a hybrid organism
d. any agricultural organism produced by breeding or biotechnology

---

Solution:

B

Exercise:

Problem:

What is the role of *Agrobacterium tumefaciens* in the production of transgenic plants?

a. Genes from *A. tumefaciens* are inserted into plant DNA to give the plant different traits.
b. Transgenic plants have been given resistance to the pest *A. tumefaciens*.
c. *A. tumefaciens* is used as a vector to move genes into plant cells.
d. Plant genes are incorporated into the genome of *Agrobacterium tumefaciens*.

Solution:

C

Free Response

Exercise:

Problem:

Today, it is possible for a diabetic patient to purchase human insulin from a pharmacist. What technology makes this possible and why is it a benefit over how things used to be?

Solution:

The human insulin comes from the gene that produces insulin in humans, which has been spliced into a bacterial genome using recombinant DNA technology. The bacterium produces the insulin, which is then purified for human use. Before there was genetically engineered human insulin, diabetics were given insulin extracted from pig pancreases, which was similar to, but not exactly like, human

insulin. Because it was not exactly like human insulin, the pig insulin caused complications in some diabetic patients.

Glossary

gene therapy
: the technique used to cure heritable diseases by replacing mutant genes with good genes

genetic testing
: identifying gene variants in an individual that may lead to a genetic disease in that individual

Genomics and Proteomics
By the end of this section, you will be able to:

- Define genomics and proteomics
- Define whole genome sequencing
- Explain different applications of genomics and proteomics

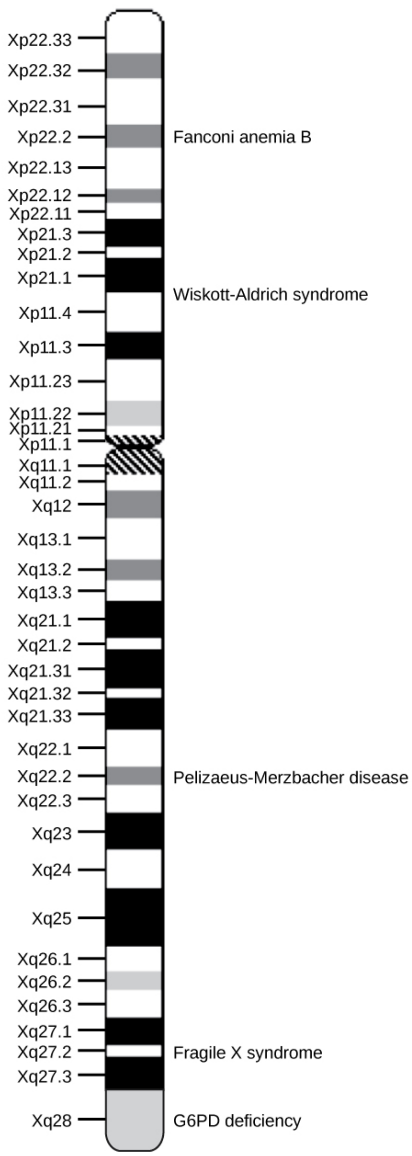
The study of nucleic acids began with the discovery of DNA, progressed to the study of genes and small fragments, and has now exploded to the field of **genomics**. Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species. The advances in genomics have been made possible by DNA sequencing technology. Just as information technology has led to Google Maps that enable us to get detailed information about locations around the globe, genomic information is used to create similar maps of the DNA of different organisms.

Mapping Genomes

Genome mapping is the process of finding the location of genes on each chromosome. The maps that are created are comparable to the maps that we use to navigate streets. A **genetic map** is an illustration that lists genes and their location on a chromosome. Genetic maps provide the big picture (similar to a map of interstate highways) and use genetic markers (similar to landmarks). A genetic marker is a gene or sequence on a chromosome that shows genetic linkage with a trait of interest. The genetic marker tends to be inherited with the gene of interest, and one measure of distance between them is the recombination frequency during meiosis. Early geneticists called this linkage analysis.

Physical maps get into the intimate details of smaller regions of the chromosomes (similar to a detailed road map) ([link]). A physical map is a representation of the physical distance, in nucleotides, between genes or genetic markers. Both genetic linkage maps and physical maps are required to build a complete picture of the genome. Having a complete map of the genome makes it easier for researchers to study individual genes. Human genome maps help researchers in their efforts to identify human disease-

causing genes related to illnesses such as cancer, heart disease, and cystic fibrosis, to name a few. In addition, genome mapping can be used to help identify organisms with beneficial traits, such as microbes with the ability to clean up pollutants or even prevent pollution. Research involving plant genome mapping may lead to methods that produce higher crop yields or to the development of plants that adapt better to climate change.



This is a physical map of the human X

chromosome. (credit: modification of work by NCBI, NIH)

Genetic maps provide the outline, and physical maps provide the details. It is easy to understand why both types of genome-mapping techniques are important to show the big picture. Information obtained from each technique is used in combination to study the genome. Genomic mapping is used with different model organisms that are used for research. Genome mapping is still an ongoing process, and as more advanced techniques are developed, more advances are expected. Genome mapping is similar to completing a complicated puzzle using every piece of available data. Mapping information generated in laboratories all over the world is entered into central databases, such as the National Center for Biotechnology Information (NCBI). Efforts are made to make the information more easily accessible to researchers and the general public. Just as we use global positioning systems instead of paper maps to navigate through roadways, NCBI allows us to use a genome viewer tool to simplify the data mining process.

Note:
Concept in Action

openstax COLLEGE™

Online Mendelian Inheritance in Man (OMIM) is a searchable online catalog of human genes and genetic disorders. This website shows genome mapping, and also details the history and research of each trait and

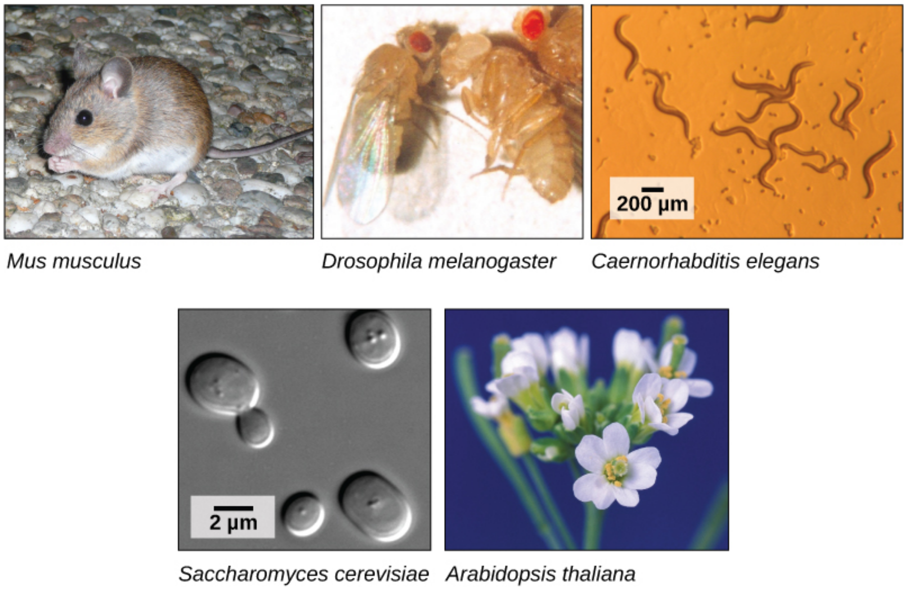
disorder. Click the link to search for traits (such as handedness) and genetic disorders (such as diabetes).

Whole Genome Sequencing

Although there have been significant advances in the medical sciences in recent years, doctors are still confounded by many diseases and researchers are using whole genome sequencing to get to the bottom of the problem. **Whole genome sequencing** is a process that determines the DNA sequence of an entire genome. Whole genome sequencing is a brute-force approach to problem solving when there is a genetic basis at the core of a disease. Several laboratories now provide services to sequence, analyze, and interpret entire genomes.

In 2010, whole genome sequencing was used to save a young boy whose intestines had multiple mysterious abscesses. The child had several colon operations with no relief. Finally, a whole genome sequence revealed a defect in a pathway that controls apoptosis (programmed cell death). A bone marrow transplant was used to overcome this genetic disorder, leading to a cure for the boy. He was the first person to be successfully diagnosed using whole genome sequencing.

The first genomes to be sequenced, such as those belonging to viruses, bacteria, and yeast, were smaller in terms of the number of nucleotides than the genomes of multicellular organisms. The genomes of other model organisms, such as the mouse (*Mus musculus*), the fruit fly (*Drosophila melanogaster*), and the nematode (*Caenorhabditis elegans*) are now known. A great deal of basic research is performed in **model organisms** because the information can be applied to other organisms. A model organism is a species that is studied as a model to understand the biological processes in other species that can be represented by the model organism. For example, fruit flies are able to metabolize alcohol like humans, so the genes affecting sensitivity to alcohol have been studied in fruit flies in an effort to understand the variation in sensitivity to alcohol in humans. Having entire genomes sequenced helps with the research efforts in these model organisms ([link]).



Much basic research is done with model organisms, such as the mouse, *Mus musculus*; the fruit fly, *Drosophila melanogaster*; the nematode *Caenorhabditis elegans*; the yeast *Saccharomyces cerevisiae*; and the common weed, *Arabidopsis thaliana*. (credit "mouse": modification of work by Florean Fortescue; credit "nematodes": modification of work by "snickclunk"/Flickr; credit "common weed": modification of work by Peggy Greb, USDA; scale-bar data from Matt Russell)

The first human genome sequence was published in 2003. The number of whole genomes that have been sequenced steadily increases and now includes hundreds of species and thousands of individual human genomes.

Applying Genomics

The introduction of DNA sequencing and whole genome sequencing projects, particularly the Human Genome Project, has expanded the applicability of DNA sequence information. Genomics is now being used in

a wide variety of fields, such as metagenomics, pharmacogenomics, and mitochondrial genomics. The most commonly known application of genomics is to understand and find cures for diseases.

Predicting Disease Risk at the Individual Level

Predicting the risk of disease involves screening and identifying currently healthy individuals by genome analysis at the individual level. Intervention with lifestyle changes and drugs can be recommended before disease onset. However, this approach is most applicable when the problem arises from a single gene mutation. Such defects only account for about 5 percent of diseases found in developed countries. Most of the common diseases, such as heart disease, are multifactorial or polygenic, which refers to a phenotypic characteristic that is determined by two or more genes, and also environmental factors such as diet. In April 2010, scientists at Stanford University published the genome analysis of a healthy individual (Stephen Quake, a scientist at Stanford University, who had his genome sequenced); the analysis predicted his propensity to acquire various diseases. A risk assessment was done to analyze Quake's percentage of risk for 55 different medical conditions. A rare genetic mutation was found that showed him to be at risk for sudden heart attack. He was also predicted to have a 23 percent risk of developing prostate cancer and a 1.4 percent risk of developing Alzheimer's disease. The scientists used databases and several publications to analyze the genomic data. Even though genomic sequencing is becoming more affordable and analytical tools are becoming more reliable, ethical issues surrounding genomic analysis at a population level remain to be addressed. For example, could such data be legitimately used to charge more or less for insurance or to affect credit ratings?

Genome-wide Association Studies

Since 2005, it has been possible to conduct a type of study called a genome-wide association study, or GWAS. A GWAS is a method that identifies differences between individuals in single nucleotide polymorphisms (SNPs)

that may be involved in causing diseases. The method is particularly suited to diseases that may be affected by one or many genetic changes throughout the genome. It is very difficult to identify the genes involved in such a disease using family history information. The GWAS method relies on a genetic database that has been in development since 2002 called the International HapMap Project. The HapMap Project sequenced the genomes of several hundred individuals from around the world and identified groups of SNPs. The groups include SNPs that are located near to each other on chromosomes so they tend to stay together through recombination. The fact that the group stays together means that identifying one marker SNP is all that is needed to identify all the SNPs in the group. There are several million SNPs identified, but identifying them in other individuals who have not had their complete genome sequenced is much easier because only the marker SNPs need to be identified.

In a common design for a GWAS, two groups of individuals are chosen; one group has the disease, and the other group does not. The individuals in each group are matched in other characteristics to reduce the effect of confounding variables causing differences between the two groups. For example, the genotypes may differ because the two groups are mostly taken from different parts of the world. Once the individuals are chosen, and typically their numbers are a thousand or more for the study to work, samples of their DNA are obtained. The DNA is analyzed using automated systems to identify large differences in the percentage of particular SNPs between the two groups. Often the study examines a million or more SNPs in the DNA. The results of GWAS can be used in two ways: the genetic differences may be used as markers for susceptibility to the disease in undiagnosed individuals, and the particular genes identified can be targets for research into the molecular pathway of the disease and potential therapies. An offshoot of the discovery of gene associations with disease has been the formation of companies that provide so-called “personal genomics” that will identify risk levels for various diseases based on an individual’s SNP complement. The science behind these services is controversial.

Because GWAS looks for associations between genes and disease, these studies provide data for other research into causes, rather than answering

specific questions themselves. An association between a gene difference and a disease does not necessarily mean there is a cause-and-effect relationship. However, some studies have provided useful information about the genetic causes of diseases. For example, three different studies in 2005 identified a gene for a protein involved in regulating inflammation in the body that is associated with a disease-causing blindness called age-related macular degeneration. This opened up new possibilities for research into the cause of this disease. A large number of genes have been identified to be associated with Crohn's disease using GWAS, and some of these have suggested new hypothetical mechanisms for the cause of the disease.

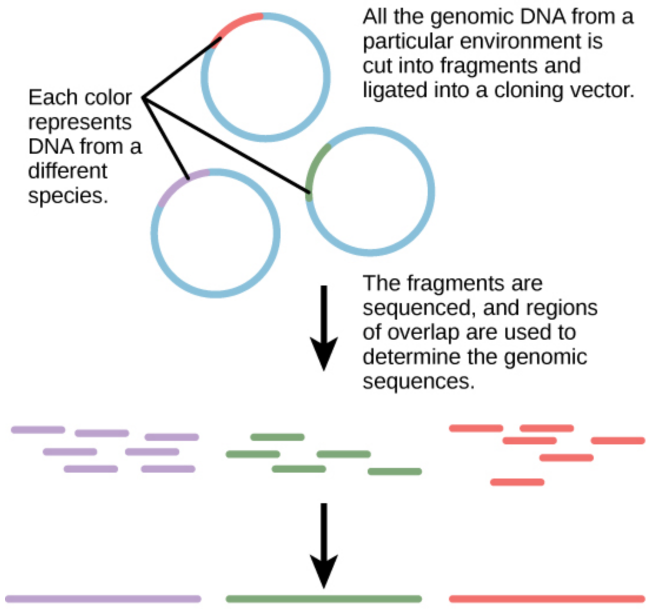
Pharmacogenomics

Pharmacogenomics involves evaluating the effectiveness and safety of drugs on the basis of information from an individual's genomic sequence. Personal genome sequence information can be used to prescribe medications that will be most effective and least toxic on the basis of the individual patient's genotype. Studying changes in gene expression could provide information about the gene transcription profile in the presence of the drug, which can be used as an early indicator of the potential for toxic effects. For example, genes involved in cellular growth and controlled cell death, when disturbed, could lead to the growth of cancerous cells. Genome-wide studies can also help to find new genes involved in drug toxicity. The gene signatures may not be completely accurate, but can be tested further before pathologic symptoms arise.

Metagenomics

Traditionally, microbiology has been taught with the view that microorganisms are best studied under pure culture conditions, which involves isolating a single type of cell and culturing it in the laboratory. Because microorganisms can go through several generations in a matter of hours, their gene expression profiles adapt to the new laboratory environment very quickly. On the other hand, many species resist being

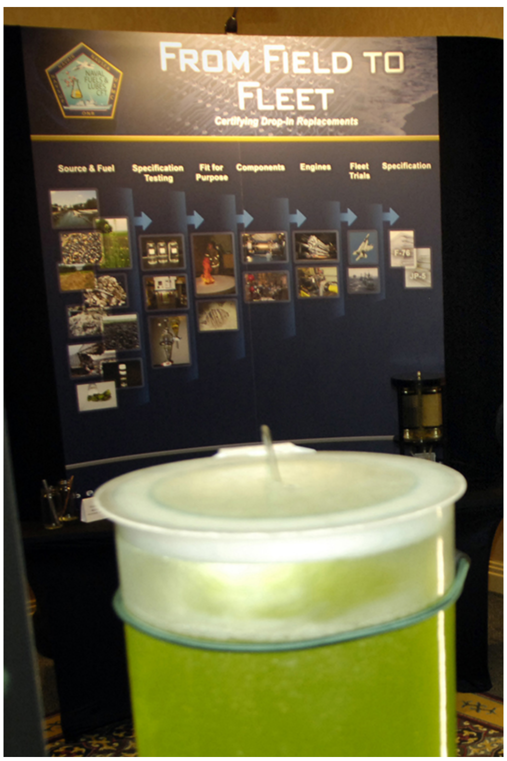
cultured in isolation. Most microorganisms do not live as isolated entities, but in microbial communities known as biofilms. For all of these reasons, pure culture is not always the best way to study microorganisms. **Metagenomics** is the study of the collective genomes of multiple species that grow and interact in an environmental niche. Metagenomics can be used to identify new species more rapidly and to analyze the effect of pollutants on the environment ([link]). Metagenomics techniques can now also be applied to communities of higher eukaryotes, such as fish.



Metagenomics involves isolating DNA from multiple species within an environmental niche. The DNA is cut up and sequenced, allowing entire genome sequences of multiple species to be reconstructed from the sequences of overlapping pieces.

Creation of New Biofuels

Knowledge of the genomics of microorganisms is being used to find better ways to harness biofuels from algae and cyanobacteria. The primary sources of fuel today are coal, oil, wood, and other plant products such as ethanol. Although plants are renewable resources, there is still a need to find more alternative renewable sources of energy to meet our population's energy demands. The microbial world is one of the largest resources for genes that encode new enzymes and produce new organic compounds, and it remains largely untapped. This vast genetic resource holds the potential to provide new sources of biofuels ([link]).



Renewable fuels were tested in Navy ships and aircraft at the first Naval Energy Forum. (credit: modification of work by

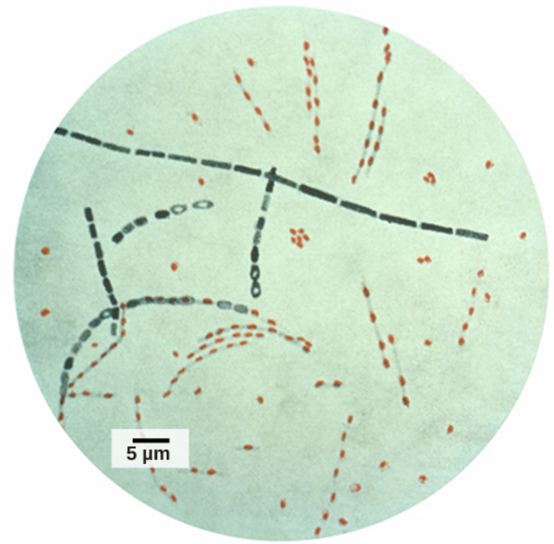
John F. Williams, US
Navy)

Mitochondrial Genomics

Mitochondria are intracellular organelles that contain their own DNA. Mitochondrial DNA mutates at a rapid rate and is often used to study evolutionary relationships. Another feature that makes studying the mitochondrial genome interesting is that in most multicellular organisms, the mitochondrial DNA is passed on from the mother during the process of fertilization. For this reason, mitochondrial genomics is often used to trace genealogy.

Genomics in Forensic Analysis

Information and clues obtained from DNA samples found at crime scenes have been used as evidence in court cases, and genetic markers have been used in forensic analysis. Genomic analysis has also become useful in this field. In 2001, the first use of genomics in forensics was published. It was a collaborative effort between academic research institutions and the FBI to solve the mysterious cases of anthrax ([link]) that was transported by the US Postal Service. Anthrax bacteria were made into an infectious powder and mailed to news media and two U.S. Senators. The powder infected the administrative staff and postal workers who opened or handled the letters. Five people died, and 17 were sickened from the bacteria. Using microbial genomics, researchers determined that a specific strain of anthrax was used in all the mailings; eventually, the source was traced to a scientist at a national biodefense laboratory in Maryland.



Bacillus anthracis is the organism that causes anthrax. (credit: modification of work by CDC; scale-bar data from Matt Russell)

Genomics in Agriculture

Genomics can reduce the trials and failures involved in scientific research to a certain extent, which could improve the quality and quantity of crop yields in agriculture ([link]). Linking traits to genes or gene signatures helps to improve crop breeding to generate hybrids with the most desirable qualities. Scientists use genomic data to identify desirable traits, and then transfer those traits to a different organism to create a new genetically modified organism, as described in the previous module. Scientists are discovering how genomics can improve the quality and quantity of agricultural production. For example, scientists could use desirable traits to create a useful product or enhance an existing product, such as making a drought-sensitive crop more tolerant of the dry season.

Transgenic agricultural plants can be made to resist disease. These transgenic plums are resistant to the plum pox virus. (credit: Scott Bauer, USDA ARS)

Proteomics

Proteins are the final products of genes that perform the function encoded by the gene. Proteins are composed of amino acids and play important roles in the cell. All enzymes (except ribozymes) are proteins and act as catalysts that affect the rate of reactions. Proteins are also regulatory molecules, and some are hormones. Transport proteins, such as hemoglobin, help transport oxygen to various organs. Antibodies that defend against foreign particles are also proteins. In the diseased state, protein function can be impaired because of changes at the genetic level or because of direct impact on a specific protein.

A proteome is the entire set of proteins produced by a cell type. Proteomes can be studied using the knowledge of genomes because genes code for mRNAs, and the mRNAs encode proteins. The study of the function of proteomes is called **proteomics**. Proteomics complements genomics and is

useful when scientists want to test their hypotheses that were based on genes. Even though all cells in a multicellular organism have the same set of genes, the set of proteins produced in different tissues is different and dependent on gene expression. Thus, the genome is constant, but the proteome varies and is dynamic within an organism. In addition, RNAs can be alternatively spliced (cut and pasted to create novel combinations and novel proteins), and many proteins are modified after translation. Although the genome provides a blueprint, the final architecture depends on several factors that can change the progression of events that generate the proteome.

Genomes and proteomes of patients suffering from specific diseases are being studied to understand the genetic basis of the disease. The most prominent disease being studied with proteomic approaches is cancer ([link]). Proteomic approaches are being used to improve the screening and early detection of cancer; this is achieved by identifying proteins whose expression is affected by the disease process. An individual protein is called a **biomarker**, whereas a set of proteins with altered expression levels is called a **protein signature**. For a biomarker or protein signature to be useful as a candidate for early screening and detection of a cancer, it must be secreted in body fluids such as sweat, blood, or urine, so that large-scale screenings can be performed in a noninvasive fashion. The current problem with using biomarkers for the early detection of cancer is the high rate of false-negative results. A false-negative result is a negative test result that should have been positive. In other words, many cases of cancer go undetected, which makes biomarkers unreliable. Some examples of protein biomarkers used in cancer detection are CA-125 for ovarian cancer and PSA for prostate cancer. Protein signatures may be more reliable than biomarkers to detect cancer cells. Proteomics is also being used to develop individualized treatment plans, which involves the prediction of whether or not an individual will respond to specific drugs and the side effects that the individual may have. Proteomics is also being used to predict the possibility of disease recurrence.

This machine is preparing to do a proteomic pattern analysis to identify specific cancers so that an accurate cancer prognosis can be made. (credit: Dorie Hightower, NCI, NIH)

The National Cancer Institute has developed programs to improve the detection and treatment of cancer. The Clinical Proteomic Technologies for Cancer and the Early Detection Research Network are efforts to identify protein signatures specific to different types of cancers. The Biomedical Proteomics Program is designed to identify protein signatures and design effective therapies for cancer patients.

Section Summary

Genome mapping is similar to solving a big, complicated puzzle with pieces of information coming from laboratories all over the world. Genetic maps provide an outline for the location of genes within a genome, and they estimate the distance between genes and genetic markers on the basis of the recombination frequency during meiosis. Physical maps provide detailed information about the physical distance between the genes. The most detailed information is available through sequence mapping. Information

from all mapping and sequencing sources is combined to study an entire genome.

Whole genome sequencing is the latest available resource to treat genetic diseases. Some doctors are using whole genome sequencing to save lives. Genomics has many industrial applications, including biofuel development, agriculture, pharmaceuticals, and pollution control.

Imagination is the only barrier to the applicability of genomics. Genomics is being applied to most fields of biology; it can be used for personalized medicine, prediction of disease risks at an individual level, the study of drug interactions before the conduction of clinical trials, and the study of microorganisms in the environment as opposed to the laboratory. It is also being applied to the generation of new biofuels, genealogical assessment using mitochondria, advances in forensic science, and improvements in agriculture.

Proteomics is the study of the entire set of proteins expressed by a given type of cell under certain environmental conditions. In a multicellular organism, different cell types will have different proteomes, and these will vary with changes in the environment. Unlike a genome, a proteome is dynamic and under constant flux, which makes it more complicated and more useful than the knowledge of genomes alone.

Multiple Choice

Exercise:

Problem:

What is the most challenging issue facing genome sequencing?

a. the inability to develop fast and accurate sequencing techniques
b. the ethics of using information from genomes at the individual level
c. the availability and stability of DNA
d. all of the above

Solution:

B

Exercise:

Problem: Genomics can be used in agriculture to:

a. generate new hybrid strains
b. improve disease resistance
c. improve yield
d. all of the above

Solution:

D

Exercise:

Problem:

What kind of diseases are studied using genome-wide association studies?

a. viral diseases
b. single-gene inherited diseases
c. diseases caused by multiple genes
d. diseases caused by environmental factors

Solution:

C

Free Response

Exercise:

Problem: Describe two of the applications for genome mapping.

Solution:

Genome mapping helps researchers to study disease-causing genes in humans. It also helps to identify traits of organisms that can be used in applications such as cleaning up pollution.

Exercise:

Problem:

Identify a possible advantage and a possible disadvantage of a genetic test that would identify genes in individuals that increase their probability of having Alzheimer's disease later in life.

Solution:

The benefit of such a test is that the individual can make preparations for having the disease including taking treatments that slow the disease. The disadvantage of the test is that it might be used by insurance companies to deny coverage to the person.

Glossary

biomarker
: an individual protein that is uniquely produced in a diseased state

genetic map
: an outline of genes and their location on a chromosome that is based on recombination frequencies between markers

genomics
: the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species

metagenomics
: the study of the collective genomes of multiple species that grow and interact in an environmental niche

model organism
: a species that is studied and used as a model to understand the biological processes in other species represented by the model organism

pharmacogenomics
: the study of drug interactions with the genome or proteome; also called toxicogenomics

physical map
: a representation of the physical distance between genes or genetic markers

protein signature
: a set of over- or under-expressed proteins characteristic of cells in a particular diseased tissue

proteomics
: study of the function of proteomes

whole genome sequencing
: a process that determines the nucleotide sequence of an entire genome

Introduction class="introduction"

The diversity of life on Earth is the result of evolution, a continuous process that is still occurring. (credit “wolf”: modification of work by Gary Kramer, USFWS; credit “coral”: modification of work by William Harrigan, NOAA; credit “river”: modification of work by Vojtěch Dostál; credit “protozoa”: modification of work by Sharon

Franklin, Stephen Ausmus, USDA ARS; credit “fish” modification of work by Christian Mehlführer; credit “mushroom”, “bee”: modification of work by Cory Zanker; credit “tree”: modification of work by Joseph Kranak)

All species of living organisms—from the bacteria on our skin, to the trees in our yards, to the birds outside—evolved at some point from a different species. Although it may seem that living things today stay much the same from generation to generation, that is not the case: evolution is ongoing. Evolution is the process through which the characteristics of species change and through which new species arise.

The theory of evolution is the unifying theory of biology, meaning it is the framework within which biologists ask questions about the living world. Its power is that it provides direction for predictions about living things that are borne out in experiment after experiment. The Ukrainian-born American geneticist Theodosius Dobzhansky famously wrote that "nothing makes sense in biology except in the light of evolution."[footnote] He meant that the principle that all life has evolved and diversified from a common ancestor is the foundation from which we understand all other questions in biology. This chapter will explain some of the mechanisms for evolutionary change and the kinds of questions that biologists can and have answered using evolutionary theory.

Theodosius Dobzhansky. “Biology, Molecular and Organismic.” *American Zoologist* 4, no. 4 (1964): 449.

Discovering How Populations Change

By the end of this section, you will be able to:

- Explain how Darwin’s theory of evolution differed from the current view at the time
- Describe how the present-day theory of evolution was developed
- Describe how population genetics is used to study the evolution of populations

The theory of evolution by natural selection describes a mechanism for species change over time. That species change had been suggested and debated well before Darwin. The view that species were static and unchanging was grounded in the writings of Plato, yet there were also ancient Greeks that expressed evolutionary ideas.

In the eighteenth century, ideas about the evolution of animals were reintroduced by the naturalist Georges-Louis Leclerc, Comte de Buffon and even by Charles Darwin’s grandfather, Erasmus Darwin. During this time, it was also accepted that there were extinct species. At the same time, James Hutton, the Scottish naturalist, proposed that geological change occurred gradually by the accumulation of small changes from processes (over long periods of time) just like those happening today. This contrasted with the predominant view that the geology of the planet was a consequence of catastrophic events occurring during a relatively brief past. Hutton’s view was later popularized by the geologist Charles Lyell in the nineteenth century. Lyell became a friend to Darwin and his ideas were very influential on Darwin’s thinking. Lyell argued that the greater age of Earth gave more time for gradual change in species, and the process provided an analogy for gradual change in species.

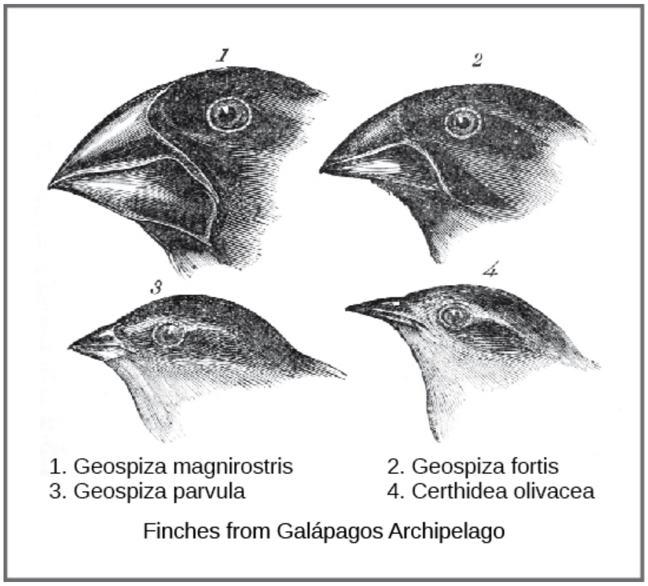
In the early nineteenth century, Jean-Baptiste Lamarck published a book that detailed a mechanism for evolutionary change that is now referred to as **inheritance of acquired characteristics**. In Lamarck’s theory, modifications in an individual caused by its environment, or the use or disuse of a structure during its lifetime, could be inherited by its offspring and, thus, bring about change in a species. While this mechanism for evolutionary change as described by Lamarck was discredited, Lamarck’s ideas were an important influence on evolutionary thought. The inscription

on the statue of Lamarck that stands at the gates of the Jardin des Plantes in Paris describes him as the "founder of the doctrine of evolution."

Charles Darwin and Natural Selection

The actual mechanism for evolution was independently conceived of and described by two naturalists, Charles Darwin and Alfred Russell Wallace, in the mid-nineteenth century. Importantly, each spent time exploring the natural world on expeditions to the tropics. From 1831 to 1836, Darwin traveled around the world on *H.M.S. Beagle*, visiting South America, Australia, and the southern tip of Africa. Wallace traveled to Brazil to collect insects in the Amazon rainforest from 1848 to 1852 and to the Malay Archipelago from 1854 to 1862. Darwin's journey, like Wallace's later journeys in the Malay Archipelago, included stops at several island chains, the last being the Galápagos Islands (west of Ecuador). On these islands, Darwin observed species of organisms on different islands that were clearly similar, yet had distinct differences. For example, the ground finches inhabiting the Galápagos Islands comprised several species that each had a unique beak shape ([link]). He observed both that these finches closely resembled another finch species on the mainland of South America and that the group of species in the Galápagos formed a graded series of beak sizes and shapes, with very small differences between the most similar. Darwin imagined that the island species might be all species modified from one original mainland species. In 1860, he wrote, "Seeing this gradation and diversity of structure in one small, intimately related group of birds, one might really fancy that from an original paucity of birds in this archipelago, one species had been taken and modified for different ends."[footnote]

Charles Darwin, *Journal of Researches into the Natural History and Geology of the Countries Visited during the Voyage of H.M.S. Beagle Round the World, under the Command of Capt. Fitz Roy, R.N*, 2nd. ed. (London: John Murray, 1860), http://www.archive.org/details/journalofresea00darw.



Darwin observed that beak shape varies among finch species. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources. This illustration shows the beak shapes for four species of ground finch: 1. *Geospiza magnirostris* (the large ground finch), 2. *G. fortis* (the medium ground finch), 3. *G. parvula* (the small tree finch), and 4. *Certhidea olivacea* (the green-warbler finch).

Wallace and Darwin both observed similar patterns in other organisms and independently conceived a mechanism to explain how and why such changes could take place. Darwin called this mechanism natural selection. **Natural selection**, Darwin argued, was an inevitable outcome of three principles that operated in nature. First, the characteristics of organisms are inherited, or passed from parent to offspring. Second, more offspring are

produced than are able to survive; in other words, resources for survival and reproduction are limited. The capacity for reproduction in all organisms outstrips the availability of resources to support their numbers. Thus, there is a competition for those resources in each generation. Both Darwin and Wallace's understanding of this principle came from reading an essay by the economist Thomas Malthus, who discussed this principle in relation to human populations. Third, offspring vary among each other in regard to their characteristics and those variations are inherited. Out of these three principles, Darwin and Wallace reasoned that offspring with inherited characteristics that allow them to best compete for limited resources will survive and have more offspring than those individuals with variations that are less able to compete. Because characteristics are inherited, these traits will be better represented in the next generation. This will lead to change in populations over generations in a process that Darwin called "descent with modification."

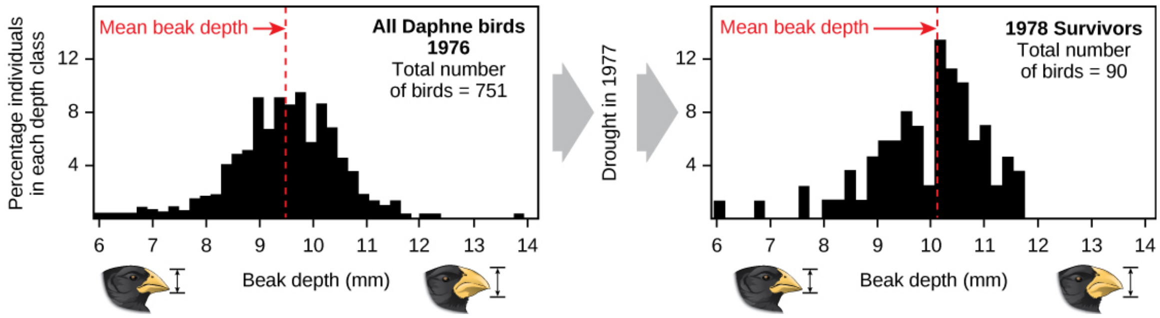
Papers by Darwin and Wallace ([link]) presenting the idea of natural selection were read together in 1858 before the Linnaean Society in London. The following year Darwin's book, *On the Origin of Species,* was published, which outlined in considerable detail his arguments for evolution by natural selection.

(a) (b)

(a) Charles Darwin and (b) Alfred Wallace wrote scientific papers on natural selection that

were presented together before the Linnean Society in 1858.

Demonstrations of evolution by natural selection can be time consuming. One of the best demonstrations has been in the very birds that helped to inspire the theory, the Galápagos finches. Peter and Rosemary Grant and their colleagues have studied Galápagos finch populations every year since 1976 and have provided important demonstrations of the operation of natural selection. The Grants found changes from one generation to the next in the beak shapes of the medium ground finches on the Galápagos island of Daphne Major. The medium ground finch feeds on seeds. The birds have inherited variation in the bill shape with some individuals having wide, deep bills and others having thinner bills. Large-billed birds feed more efficiently on large, hard seeds, whereas smaller billed birds feed more efficiently on small, soft seeds. During 1977, a drought period altered vegetation on the island. After this period, the number of seeds declined dramatically: the decline in small, soft seeds was greater than the decline in large, hard seeds. The large-billed birds were able to survive better than the small-billed birds the following year. The year following the drought when the Grants measured beak sizes in the much-reduced population, they found that the average bill size was larger ([link]). This was clear evidence for natural selection (differences in survival) of bill size caused by the availability of seeds. The Grants had studied the inheritance of bill sizes and knew that the surviving large-billed birds would tend to produce offspring with larger bills, so the selection would lead to evolution of bill size. Subsequent studies by the Grants have demonstrated selection on and evolution of bill size in this species in response to changing conditions on the island. The evolution has occurred both to larger bills, as in this case, and to smaller bills when large seeds became rare.



A drought on the Galápagos island of Daphne Major in 1977 reduced the number of small seeds available to finches, causing many of the small-beaked finches to die. This caused an increase in the finches’ average beak size between 1976 and 1978.

Variation and Adaptation

Natural selection can only take place if there is **variation**, or differences, among individuals in a population. Importantly, these differences must have some genetic basis; otherwise, selection will not lead to change in the next generation. This is critical because variation among individuals can be caused by non-genetic reasons, such as an individual being taller because of better nutrition rather than different genes.

Genetic diversity in a population comes from two main sources: mutation and sexual reproduction. Mutation, a change in DNA, is the ultimate source of new alleles or new genetic variation in any population. An individual that has a mutated gene might have a different trait than other individuals in the population. However, this is not always the case. A mutation can have one of three outcomes on the organisms’ appearance (or phenotype):

- A mutation may affect the phenotype of the organism in a way that gives it reduced fitness—lower likelihood of survival, resulting in fewer offspring.
- A mutation may produce a phenotype with a beneficial effect on fitness.

- Many mutations, called neutral mutations, will have no effect on fitness.

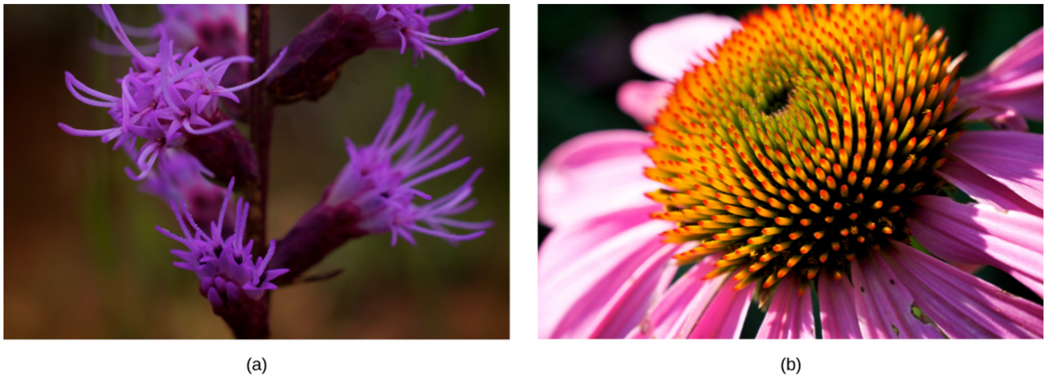
Mutations may also have a whole range of effect sizes on the fitness of the organism that expresses them in their phenotype, from a small effect to a great effect. Sexual reproduction and crossing over in meiosis also lead to genetic diversity: when two parents reproduce, unique combinations of alleles assemble to produce unique genotypes and, thus, phenotypes in each of the offspring.

A heritable trait that aids the survival and reproduction of an organism in its present environment is called an **adaptation**. An adaptation is a “match” of the organism to the environment. Adaptation to an environment comes about when a change in the range of genetic variation occurs over time that increases or maintains the match of the population with its environment. The variations in finch beaks shifted from generation to generation providing adaptation to food availability.

Whether or not a trait is favorable depends on the environment at the time. The same traits do not always have the same relative benefit or disadvantage because environmental conditions can change. For example, finches with large bills were benefited in one climate, while small bills were a disadvantage; in a different climate, the relationship reversed.

Patterns of Evolution

The evolution of species has resulted in enormous variation in form and function. When two species evolve in different directions from a common point, it is called **divergent evolution**. Such divergent evolution can be seen in the forms of the reproductive organs of flowering plants, which share the same basic anatomies; however, they can look very different as a result of selection in different physical environments, and adaptation to different kinds of pollinators ([link]).



Flowering plants evolved from a common ancestor. Notice that the (a) dense blazing star and (b) purple coneflower vary in appearance, yet both share a similar basic morphology. (credit a, b: modification of work by Cory Zanker)

In other cases, similar phenotypes evolve independently in distantly related species. For example, flight has evolved in both bats and insects, and they both have structures we refer to as wings, which are adaptations to flight. The wings of bats and insects, however, evolved from very different original structures. When similar structures arise through evolution independently in different species it is called **convergent evolution**. The wings of bats and insects are called **analogous structures**; they are similar in function and appearance, but do not share an origin in a common ancestor. Instead they evolved independently in the two lineages. The wings of a hummingbird and an ostrich are **homologous structures**, meaning they share similarities (despite their differences resulting from evolutionary divergence). The wings of hummingbirds and ostriches did not evolve independently in the hummingbird lineage and the ostrich lineage—they descended from a common ancestor with wings.

The Modern Synthesis

The mechanisms of inheritance, genetics, were not understood at the time Darwin and Wallace were developing their idea of natural selection. This lack of understanding was a stumbling block to comprehending many

aspects of evolution. In fact, blending inheritance was the predominant (and incorrect) genetic theory of the time, which made it difficult to understand how natural selection might operate. Darwin and Wallace were unaware of the genetics work by Austrian monk Gregor Mendel, which was published in 1866, not long after publication of *On the Origin of Species*. Mendel's work was rediscovered in the early twentieth century at which time geneticists were rapidly coming to an understanding of the basics of inheritance. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades genetics and evolution were integrated in what became known as the **modern synthesis**—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today. In sum, the modern synthesis describes how evolutionary pressures, such as natural selection, can affect a population's genetic makeup, and, in turn, how this can result in the gradual evolution of populations and species. The theory also connects this gradual change of a population over time, called **microevolution**, with the processes that gave rise to new species and higher taxonomic groups with widely divergent characters, called **macroevolution**.

Population Genetics

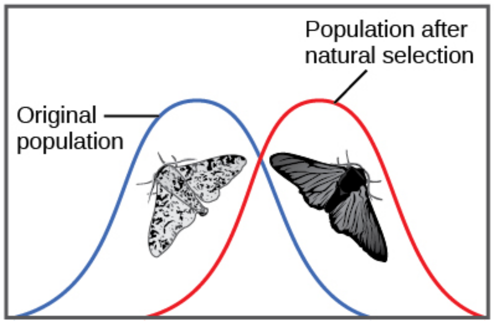
Recall that a gene for a particular character may have several variants, or alleles, that code for different traits associated with that character. For example, in the ABO blood type system in humans, three alleles determine the particular blood-type protein on the surface of red blood cells. Each individual in a population of diploid organisms can only carry two alleles for a particular gene, but more than two may be present in the individuals that make up the population. Mendel followed alleles as they were inherited from parent to offspring. In the early twentieth century, biologists began to study what happens to all the alleles in a population in a field of study known as **population genetics**.

Until now, we have defined evolution as a change in the characteristics of a population of organisms, but behind that phenotypic change is genetic change. In population genetic terms, evolution is defined as a change in the frequency of an allele in a population. Using the ABO system as an

example, the frequency of one of the alleles, I^A, is the number of copies of that allele divided by all the copies of the ABO gene in the population. For example, a study in Jordan found a frequency of I^A to be 26.1 percent. [footnote] The I^B, I^0 alleles made up 13.4 percent and 60.5 percent of the alleles respectively, and all of the frequencies add up to 100 percent. A change in this frequency over time would constitute evolution in the population.

Sahar S. Hanania, Dhia S. Hassawi, and Nidal M. Irshaid, "Allele Frequency and Molecular Genotypes of ABO Blood Group System in a Jordanian Population," *Journal of Medical Sciences* 7 (2007): 51-58, doi:10.3923/jms.2007.51.58

There are several ways the allele frequencies of a population can change. One of those ways is natural selection. If a given allele confers a phenotype that allows an individual to have more offspring that survive and reproduce, that allele, by virtue of being inherited by those offspring, will be in greater frequency in the next generation. Since allele frequencies always add up to 100 percent, an increase in the frequency of one allele always means a corresponding decrease in one or more of the other alleles. Highly beneficial alleles may, over a very few generations, become "fixed" in this way, meaning that every individual of the population will carry the allele. Similarly, detrimental alleles may be swiftly eliminated from the **gene pool**, the sum of all the alleles in a population. Part of the study of population genetics is tracking how selective forces change the allele frequencies in a population over time, which can give scientists clues regarding the selective forces that may be operating on a given population. The studies of changes in wing coloration in the peppered moth from mottled white to dark in response to soot-covered tree trunks and then back to mottled white when factories stopped producing so much soot is a classic example of studying evolution in natural populations ([link]).



Light-colored peppered moths are better camouflaged against a pristine environment; likewise, dark-colored peppered moths are better camouflaged against a sooty environment. Thus, as the Industrial Revolution progressed in nineteenth-century England, the color of the moth population shifted from light to dark.

As the Industrial Revolution caused trees to darken from soot, darker colored peppered moths were better camouflaged than the lighter colored ones, which caused there to be more of the darker colored moths in the population.

In the early twentieth century, English mathematician Godfrey Hardy and German physician Wilhelm Weinberg independently provided an explanation for a somewhat counterintuitive concept. Hardy's original explanation was in response to a misunderstanding as to why a "dominant" allele, one that masks a recessive allele, should not increase in frequency in a population until it eliminated all the other alleles. The question resulted from a common confusion about what "dominant" means, but it forced Hardy, who was not even a biologist, to point out that if there are no factors that affect an allele frequency those frequencies will remain constant from one generation to the next. This principle is now known as the Hardy-Weinberg equilibrium. The theory states that a population's allele and genotype frequencies are inherently stable—unless some kind of evolutionary force is acting on the population, the population would carry the same alleles in the same proportions generation after generation. Individuals would, as a whole, look essentially the same and this would be unrelated to whether the alleles were dominant or recessive. The four most important evolutionary forces, which will disrupt the equilibrium, are natural selection, mutation, **genetic drift**, and **migration** into or out of a population. A fifth factor, nonrandom mating, will also disrupt the Hardy-Weinberg equilibrium but only by shifting genotype frequencies, not allele frequencies. In nonrandom mating, individuals are more likely to mate with

like individuals (or unlike individuals) rather than at random. Since nonrandom mating does not change allele frequencies, it does not cause evolution directly. Natural selection has been described. Mutation creates one allele out of another one and changes an allele’s frequency by a small, but continuous amount each generation. Each allele is generated by a low, constant mutation rate that will slowly increase the allele’s frequency in a population if no other forces act on the allele. If natural selection acts against the allele, it will be removed from the population at a low rate leading to a frequency that results from a balance between selection and mutation. This is one reason that genetic diseases remain in the human population at very low frequencies. If the allele is favored by selection, it will increase in frequency. Genetic drift causes random changes in allele frequencies when populations are small. Genetic drift can often be important in evolution, as discussed in the next section. Finally, if two populations of a species have different allele frequencies, migration of individuals between them will cause frequency changes in both populations. As it happens, there is no population in which one or more of these processes are not operating, so populations are always evolving, and the Hardy-Weinberg equilibrium will never be exactly observed. However, the Hardy-Weinberg principle gives scientists a baseline expectation for allele frequencies in a non-evolving population to which they can compare evolving populations and thereby infer what evolutionary forces might be at play. The population is evolving if the frequencies of alleles or genotypes deviate from the value expected from the Hardy-Weinberg principle.

Darwin identified a special case of natural selection that he called sexual selection. Sexual selection affects an individual’s ability to mate and thus produce offspring, and it leads to the evolution of dramatic traits that often appear maladaptive in terms of survival but persist because they give their owners greater reproductive success. Sexual selection occurs in two ways: through male–male competition for mates and through female selection of mates. Male–male competition takes the form of conflicts between males, which are often ritualized, but may also pose significant threats to a male’s survival. Sometimes the competition is for territory, with females more likely to mate with males with higher quality territories. Female choice occurs when females choose a male based on a particular trait, such as feather colors, the performance of a mating dance, or the building of an

elaborate structure. In some cases male–male competition and female choice combine in the mating process. In each of these cases, the traits selected for, such as fighting ability or feather color and length, become enhanced in the males. In general, it is thought that sexual selection can proceed to a point at which natural selection against a character's further enhancement prevents its further evolution because it negatively impacts the male's ability to survive. For example, colorful feathers or an elaborate display make the male more obvious to predators.

Section Summary

Evolution by natural selection arises from three conditions: individuals within a species vary, some of those variations are heritable, and organisms have more offspring than resources can support. The consequence is that individuals with relatively advantageous variations will be more likely to survive and have higher reproductive rates than those individuals with different traits. The advantageous traits will be passed on to offspring in greater proportion. Thus, the trait will have higher representation in the next and subsequent generations leading to genetic change in the population.

The modern synthesis of evolutionary theory grew out of the reconciliation of Darwin's, Wallace's, and Mendel's thoughts on evolution and heredity. Population genetics is a theoretical framework for describing evolutionary change in populations through the change in allele frequencies. Population genetics defines evolution as a change in allele frequency over generations. In the absence of evolutionary forces allele frequencies will not change in a population; this is known as Hardy-Weinberg equilibrium principle. However, in all populations, mutation, natural selection, genetic drift, and migration act to change allele frequencies.

Multiple Choice

Exercise:

Problem:

Which scientific concept did Charles Darwin and Alfred Wallace independently discover?

a. mutation
b. natural selection
c. overbreeding
d. sexual reproduction

Solution:

B

Exercise:

Problem:

Which of the following situations will lead to natural selection?

a. The seeds of two plants land near each other and one grows larger than the other.
b. Two types of fish eat the same kind of food, and one is better able to gather food than the other.
c. Male lions compete for the right to mate with females, with only one possible winner.
d. all of the above

Solution:

D

Exercise:

Problem: What is the difference between micro- and macroevolution?

a. Microevolution describes the evolution of small organisms, such as insects, while macroevolution describes the evolution of large organisms, like people and elephants.
b. Microevolution describes the evolution of microscopic entities, such as molecules and proteins, while macroevolution describes the evolution of whole organisms.
c. Microevolution describes the evolution of populations, while macroevolution describes the emergence of new species over long periods of time.
d. Microevolution describes the evolution of organisms over their lifetimes, while macroevolution describes the evolution of organisms over multiple generations.

Solution:

C

Exercise:

Problem: Population genetics is the study of _________.

a. how allele frequencies in a population change over time
b. populations of cells in an individual
c. the rate of population growth
d. how genes affect embryological development

Solution:

A

Free Response

Exercise:

Problem:

If a person scatters a handful of plant seeds from one species in an area, how would natural selection work in this situation?

Solution:

The plants that can best use the resources of the area, including competing with other individuals for those resources, will produce more seeds themselves and those traits that allowed them to better use the resources will increase in the population of the next generation.

Exercise:

Problem: Explain the Hardy-Weinberg principle of equilibrium.

Solution:

The Hardy-Weinberg principle of equilibrium states that a population's allele frequencies are inherently stable. Unless an evolutionary force is acting upon the population, the population would carry the same genes at the same frequencies generation after generation, and individuals would, as a whole, look essentially the same.

Glossary

adaptation
: a heritable trait or behavior in an organism that aids in its survival in its present environment

analogous structure
: a structure that is similar because of evolution in response to similar selection pressures resulting in convergent evolution, not similar because of descent from a common ancestor

convergent evolution
: an evolution that results in similar forms on different species

divergent evolution
an evolution that results in different forms in two species with a common ancestor

gene pool
all of the alleles carried by all of the individuals in the population

genetic drift
the effect of chance on a population’s gene pool

homologous structure
a structure that is similar because of descent from a common ancestor

inheritance of acquired characteristics
a phrase that describes the mechanism of evolution proposed by Lamarck in which traits acquired by individuals through use or disuse could be passed on to their offspring thus leading to evolutionary change in the population

macroevolution
a broader scale of evolutionary changes seen over paleontological time

microevolution
the changes in a population’s genetic structure (i.e., allele frequency)

migration
the movement of individuals of a population to a new location; in population genetics it refers to the movement of individuals and their alleles from one population to another, potentially changing allele frequencies in both the old and the new population

modern synthesis
the overarching evolutionary paradigm that took shape by the 1940s and is generally accepted today

natural selection
the greater relative survival and reproduction of individuals in a population that have favorable heritable traits, leading to evolutionary

change

population genetics

the study of how selective forces change the allele frequencies in a population over time

variation

the variety of alleles in a population

Mechanisms of Evolution
By the end of this section, you will be able to:

- Describe the four basic causes of evolution: natural selection, mutation, genetic drift, and gene flow
- Explain how each evolutionary force can influence the allele frequencies of a population

The Hardy-Weinberg equilibrium principle says that allele frequencies in a population will remain constant in the absence of the four factors that could change them. Those factors are natural selection, mutation, genetic drift, and migration (gene flow). In fact, we know they are probably always affecting populations.

Natural Selection

Natural selection has already been discussed. Alleles are expressed in a phenotype. Depending on the environmental conditions, the phenotype confers an advantage or disadvantage to the individual with the phenotype relative to the other phenotypes in the population. If it is an advantage, then that individual will likely have more offspring than individuals with the other phenotypes, and this will mean that the allele behind the phenotype will have greater representation in the next generation. If conditions remain the same, those offspring, which are carrying the same allele, will also benefit. Over time, the allele will increase in frequency in the population.

Mutation

Mutation is a source of new alleles in a population. Mutation is a change in the DNA sequence of the gene. A mutation can change one allele into another, but the net effect is a change in frequency. The change in frequency resulting from mutation is small, so its effect on evolution is small unless it interacts with one of the other factors, such as selection. A mutation may produce an allele that is selected against, selected for, or selectively neutral. Harmful mutations are removed from the population by selection and will generally only be found in very low frequencies equal to the mutation rate. Beneficial mutations will spread through the population through selection,

although that initial spread is slow. Whether or not a mutation is beneficial or harmful is determined by whether it helps an organism survive to sexual maturity and reproduce. It should be noted that mutation is the ultimate source of genetic variation in all populations—new alleles, and, therefore, new genetic variations arise through mutation.

Genetic Drift

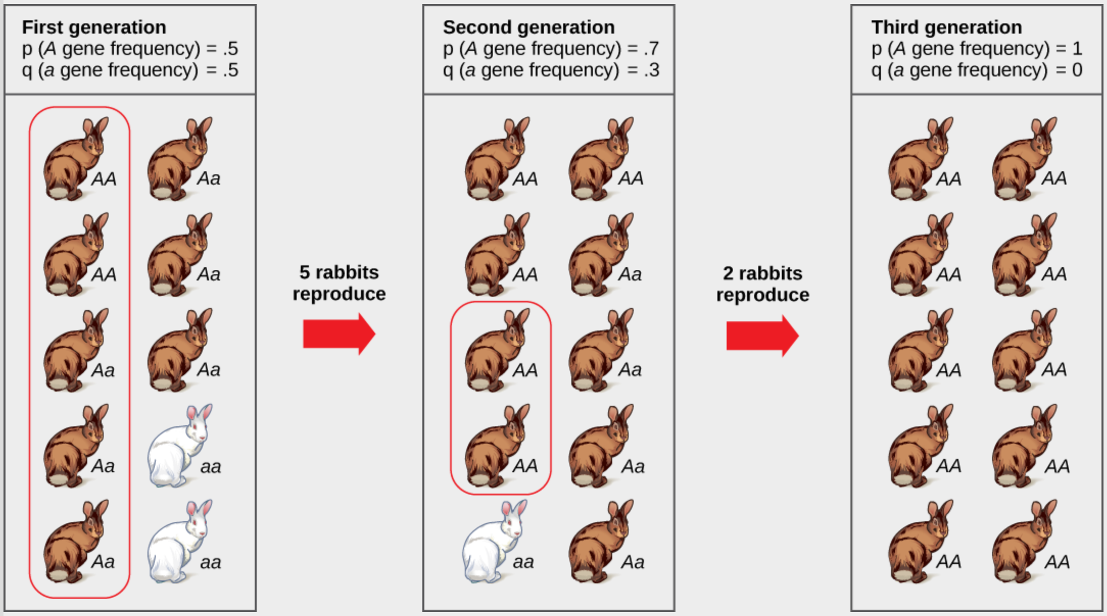
Another way a population's allele frequencies can change is genetic drift ([link]), which is simply the effect of chance. Genetic drift is most important in small populations. Drift would be completely absent in a population with infinite individuals, but, of course, no population is this large. Genetic drift occurs because the alleles in an offspring generation are a random sample of the alleles in the parent generation. Alleles may or may not make it into the next generation due to chance events including mortality of an individual, events affecting finding a mate, and even the events affecting which gametes end up in fertilizations. If one individual in a population of ten individuals happens to die before it leaves any offspring to the next generation, all of its genes—a tenth of the population's gene pool—will be suddenly lost. In a population of 100, that 1 individual represents only 1 percent of the overall gene pool; therefore, it has much less impact on the population's genetic structure and is unlikely to remove all copies of even a relatively rare allele.

Imagine a population of ten individuals, half with allele *A* and half with allele *a* (the individuals are haploid). In a stable population, the next generation will also have ten individuals. Choose that generation randomly by flipping a coin ten times and let heads be *A* and tails be *a*. It is unlikely that the next generation will have exactly half of each allele. There might be six of one and four of the other, or some different set of frequencies. Thus, the allele frequencies have changed and evolution has occurred. A coin will no longer work to choose the next generation (because the odds are no longer one half for each allele). The frequency in each generation will drift up and down on what is known as a random walk until at one point either all *A* or all *a* are chosen and that allele is fixed from that point on. This could take a very long time for a large population. This simplification is not very biological, but it can be shown that real populations behave this way.

The effect of drift on frequencies is greater the smaller a population is. Its effect is also greater on an allele with a frequency far from one half. Drift will influence every allele, even those that are being naturally selected.

Note:

Art Connection

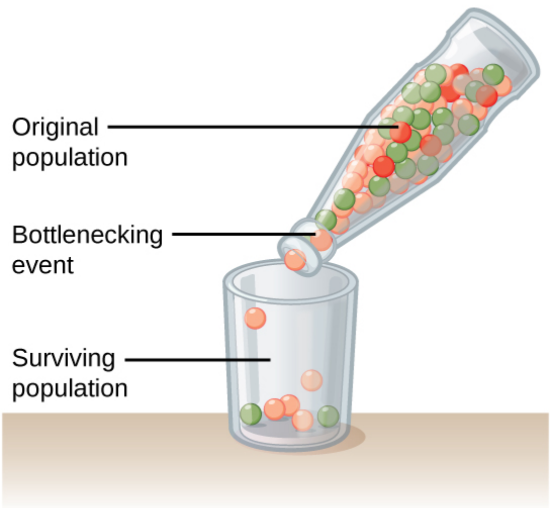


Genetic drift in a population can lead to the elimination of an allele from a population by chance. In each generation, a random set of individuals reproduces to produce the next generation. The frequency of alleles in the next generation is equal to the frequency of alleles among the individuals reproducing.

Do you think genetic drift would happen more quickly on an island or on the mainland?

Genetic drift can also be magnified by natural or human-caused events, such as a disaster that randomly kills a large portion of the population,

which is known as the **bottleneck effect** that results in a large portion of the genome suddenly being wiped out ([link]). In one fell swoop, the genetic structure of the survivors becomes the genetic structure of the entire population, which may be very different from the pre-disaster population. The disaster must be one that kills for reasons unrelated to the organism's traits, such as a hurricane or lava flow. A mass killing caused by unusually cold temperatures at night, is likely to affect individuals differently depending on the alleles they possess that confer cold hardiness.



A chance event or catastrophe can reduce the genetic variability within a population.

Another scenario in which populations might experience a strong influence of genetic drift is if some portion of the population leaves to start a new population in a new location, or if a population gets divided by a physical barrier of some kind. In this situation, those individuals are unlikely to be representative of the entire population which results in the **founder effect**. The founder effect occurs when the genetic structure matches that of the new population's founding fathers and mothers. The founder effect is

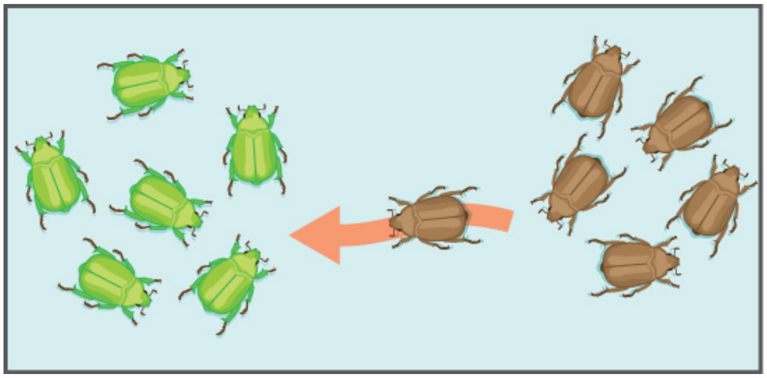
believed to have been a key factor in the genetic history of the Afrikaner population of Dutch settlers in South Africa, as evidenced by mutations that are common in Afrikaners but rare in most other populations. This is likely due to a higher-than-normal proportion of the founding colonists, which were a small sample of the original population, carried these mutations. As a result, the population expresses unusually high incidences of Huntington’s disease (HD) and Fanconi anemia (FA), a genetic disorder known to cause bone marrow and congenital abnormalities, and even cancer.[footnote]
A. J. Tipping et al., “Molecular and Genealogical Evidence for a Founder Effect in Fanconi Anemia Families of the Afrikaner Population of South Africa,” *PNAS* 98, no. 10 (2001): 5734-5739, doi: 10.1073/pnas.091402398.

Note:
Concept in Action

Visit this site to learn more about genetic drift and to run simulations of allele changes caused by drift.

Gene Flow

Another important evolutionary force is **gene flow**, or the flow of alleles in and out of a population resulting from the migration of individuals or gametes ([link]). While some populations are fairly stable, others experience more flux. Many plants, for example, send their seeds far and wide, by wind or in the guts of animals; these seeds may introduce alleles common in the source population to a new population in which they are rare.



Gene flow can occur when an individual travels from one geographic location to another and joins a different population of the species. In the example shown here, the brown allele is introduced into the green population.

Section Summary

There are four factors that can change the allele frequencies of a population. Natural selection works by selecting for alleles that confer beneficial traits or behaviors, while selecting against those for deleterious qualities. Mutations introduce new alleles into a population. Genetic drift stems from the chance occurrence that some individuals have more offspring than others and results in changes in allele frequencies that are random in direction. When individuals leave or join the population, allele frequencies can change as a result of gene flow.

Art Connections

Exercise:

Problem:

[link] Do you think genetic drift would happen more quickly on an island or on the mainland?

Solution:

[link] Genetic drift is likely to occur more rapidly on an island, where smaller populations are expected to occur.

Multiple Choice

Exercise:

Problem:

Galápagos medium ground finches are found on Santa Cruz and San Cristóbal islands, which are separated by about 100 km of ocean. Occasionally, individuals from either island fly to the other island to stay. This can alter the allele frequencies of the population through which of the following mechanisms?

a. natural selection
b. genetic drift
c. gene flow
d. mutation

Solution:

C

Exercise:

Problem:

In which of the following pairs do both evolutionary processes introduce new genetic variation into a population?

a. natural selection and genetic drift
b. mutation and gene flow
c. natural selection and gene flow
d. gene flow and genetic drift

Solution:

B

Free Response

Exercise:

Problem:

Describe natural selection and give an example of natural selection at work in a population.

Solution:

The theory of natural selection stems from the observation that some individuals in a population survive longer and have more offspring than others, thus passing on more of their genes to the next generation. For example, a big, powerful male gorilla is much more likely than a smaller, weaker gorilla to become the population's silverback, the pack's leader who mates far more than the other males of the group. The pack leader will, therefore, father more offspring, who share half of his genes, and are thus likely to also grow bigger and stronger like their father. Over time, the genes for bigger size will increase in frequency in the population, and the population will, as a result, grow larger on average.

Glossary

bottleneck effect
: the magnification of genetic drift as a result of natural events or catastrophes

founder effect

a magnification of genetic drift in a small population that migrates away from a large parent population carrying with it an unrepresentative set of alleles

gene flow
the flow of alleles in and out of a population due to the migration of individuals or gametes

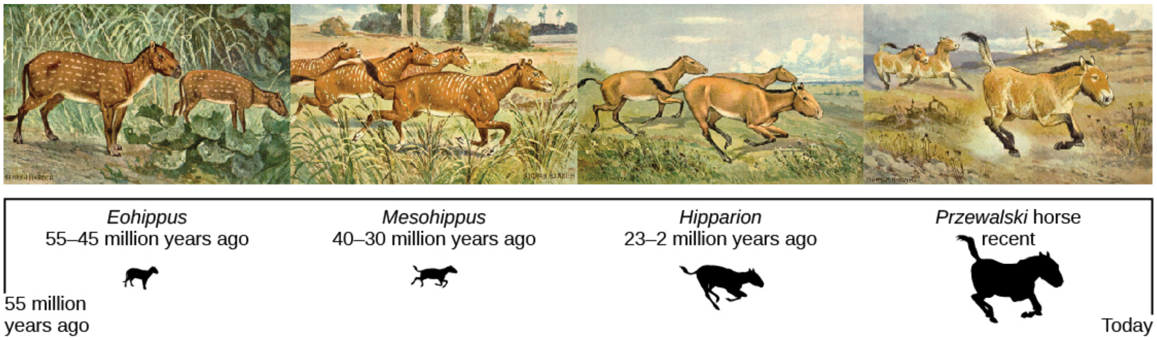
Evidence of Evolution
By the end of this section, you will be able to:

- Explain sources of evidence for evolution
- Define homologous and vestigial structures

The evidence for evolution is compelling and extensive. Looking at every level of organization in living systems, biologists see the signature of past and present evolution. Darwin dedicated a large portion of his book, *On the Origin of Species*, identifying patterns in nature that were consistent with evolution and since Darwin our understanding has become clearer and broader.

Fossils

Fossils provide solid evidence that organisms from the past are not the same as those found today; fossils show a progression of evolution. Scientists determine the age of fossils and categorize them all over the world to determine when the organisms lived relative to each other. The resulting fossil record tells the story of the past, and shows the evolution of form over millions of years ([link]). For example, highly detailed fossil records have been recovered for sequences of species in the evolution of whales and modern horses. The fossil record of horses in North America is especially rich and many contain transition fossils: those showing intermediate anatomy between earlier and later forms. The fossil record extends back to a dog-like ancestor some 55 million years ago that gave rise to the first horse-like species 55 to 42 million years ago in the genus *Eohippus*. The series of fossils tracks the change in anatomy resulting from a gradual drying trend that changed the landscape from a forested one to a prairie. Successive fossils show the evolution of teeth shapes and foot and leg anatomy to a grazing habit, with adaptations for escaping predators, for example in species of *Mesohippus* found from 40 to 30 million years ago. Later species showed gains in size, such as those of *Hipparion*, which existed from about 23 to 2 million years ago. The fossil record shows several adaptive radiations in the horse lineage, which is now much reduced to only one genus, *Equus*, with several species.

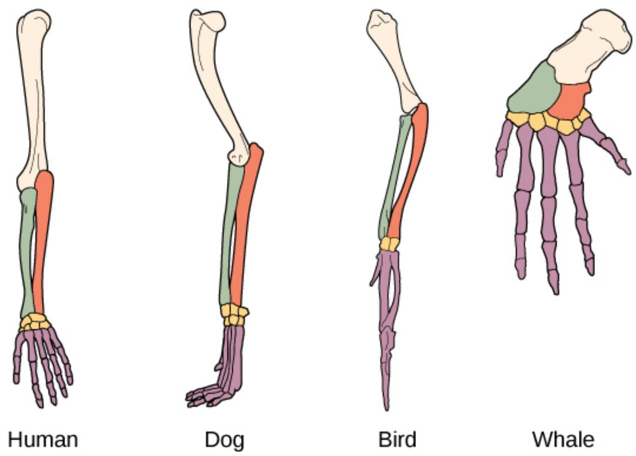


This illustration shows an artist’s renderings of these species derived from fossils of the evolutionary history of the horse and its ancestors. The species depicted are only four from a very diverse lineage that contains many branches, dead ends, and adaptive radiations. One of the trends, depicted here is the evolutionary tracking of a drying climate and increase in prairie versus forest habitat reflected in forms that are more adapted to grazing and predator escape through running. Przewalski's horse is one of a few living species of horse.

Anatomy and Embryology

Another type of evidence for evolution is the presence of structures in organisms that share the same basic form. For example, the bones in the appendages of a human, dog, bird, and whale all share the same overall construction ([link]). That similarity results from their origin in the appendages of a common ancestor. Over time, evolution led to changes in the shapes and sizes of these bones in different species, but they have maintained the same overall layout, evidence of descent from a common ancestor. Scientists call these synonymous parts homologous structures. Some structures exist in organisms that have no apparent function at all, and appear to be residual parts from a past ancestor. For example, some snakes have pelvic bones despite having no legs because they descended from reptiles that did have legs. These unused structures without function are called **vestigial structures**. Other examples of vestigial structures are wings

on flightless birds (which may have other functions), leaves on some cacti, traces of pelvic bones in whales, and the sightless eyes of cave animals.

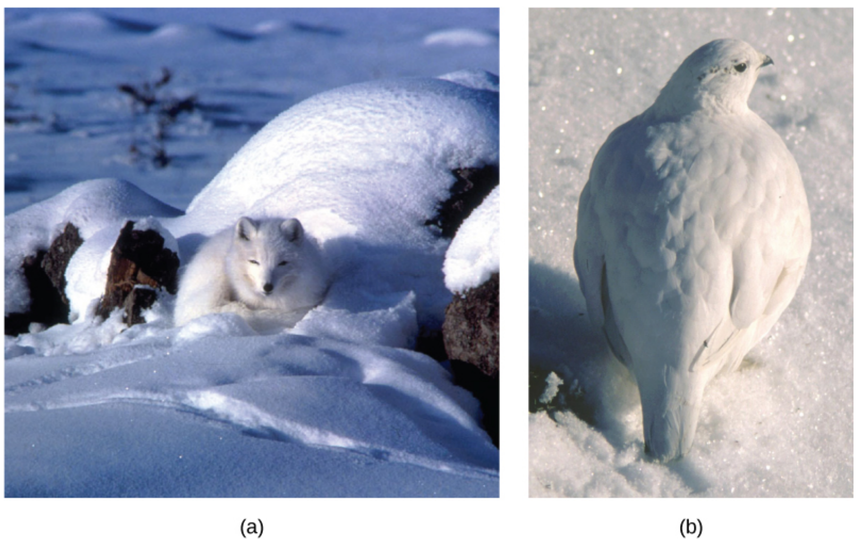


The similar construction of these appendages indicates that these organisms share a common ancestor.

Note:
Concept in Action

Click through the activities at this [interactive site]() to guess which bone structures are homologous and which are analogous, and to see examples of all kinds of evolutionary adaptations that illustrate these concepts.

Another evidence of evolution is the convergence of form in organisms that share similar environments. For example, species of unrelated animals, such as the arctic fox and ptarmigan (a bird), living in the arctic region have temporary white coverings during winter to blend with the snow and ice ([link]). The similarity occurs not because of common ancestry, indeed one covering is of fur and the other of feathers, but because of similar selection pressures—the benefits of not being seen by predators.



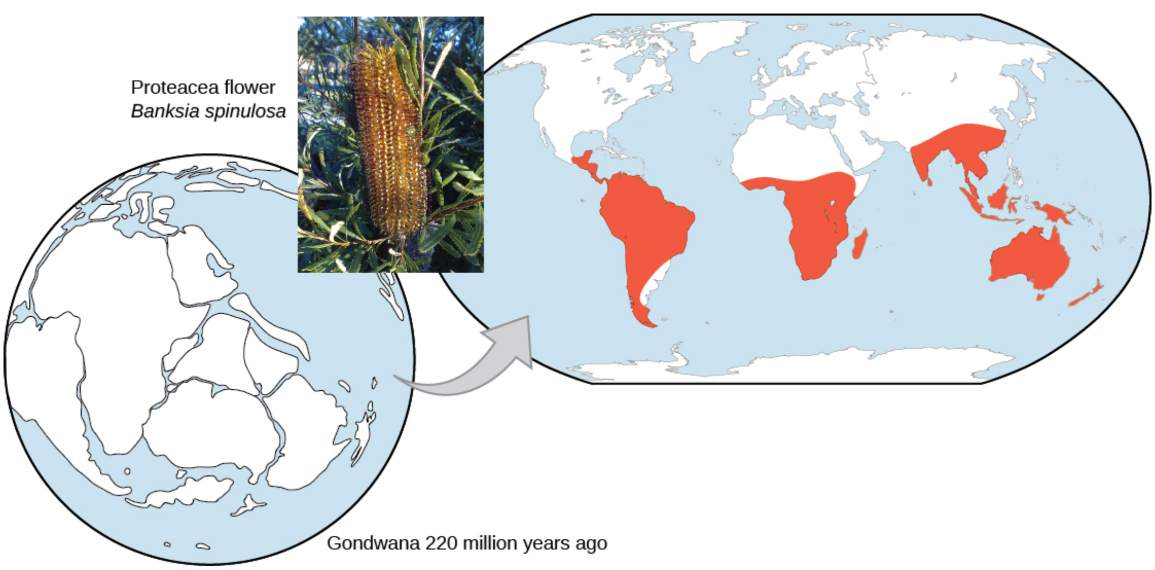
The white winter coat of (a) the arctic fox and (b) the ptarmigan's plumage are adaptations to their environments. (credit a: modification of work by Keith Morehouse)

Embryology, the study of the development of the anatomy of an organism to its adult form also provides evidence of relatedness between now widely divergent groups of organisms. Structures that are absent in some groups often appear in their embryonic forms and disappear by the time the adult or juvenile form is reached. For example, all vertebrate embryos, including humans, exhibit gill slits at some point in their early development. These disappear in the adults of terrestrial groups, but are maintained in adult forms of aquatic groups such as fish and some amphibians. Great ape

embryos, including humans, have a tail structure during their development that is lost by the time of birth. The reason embryos of unrelated species are often similar is that mutational changes that affect the organism during embryonic development can cause amplified differences in the adult, even while the embryonic similarities are preserved.

Biogeography

The geographic distribution of organisms on the planet follows patterns that are best explained by evolution in conjunction with the movement of tectonic plates over geological time. Broad groups that evolved before the breakup of the supercontinent Pangaea (about 200 million years ago) are distributed worldwide. Groups that evolved since the breakup appear uniquely in regions of the planet, for example the unique flora and fauna of northern continents that formed from the supercontinent Laurasia and of the southern continents that formed from the supercontinent Gondwana. The presence of Proteaceae in Australia, southern Africa, and South America is best explained by the plant family's presence there prior to the southern supercontinent Gondwana breaking up ([link]).



The Proteacea family of plants evolved before the supercontinent Gondwana broke up. Today, members of this plant family are

found throughout the southern hemisphere (shown in red). (credit "Proteacea flower": modification of work by "dorofofoto"/Flickr)

The great diversification of the marsupials in Australia and the absence of other mammals reflects that island continent's long isolation. Australia has an abundance of endemic species—species found nowhere else—which is typical of islands whose isolation by expanses of water prevents migration of species to other regions. Over time, these species diverge evolutionarily into new species that look very different from their ancestors that may exist on the mainland. The marsupials of Australia, the finches on the Galápagos, and many species on the Hawaiian Islands are all found nowhere else but on their island, yet display distant relationships to ancestral species on mainlands.

Molecular Biology

Like anatomical structures, the structures of the molecules of life reflect descent with modification. Evidence of a common ancestor for all of life is reflected in the universality of DNA as the genetic material and of the near universality of the genetic code and the machinery of DNA replication and expression. Fundamental divisions in life between the three domains are reflected in major structural differences in otherwise conservative structures such as the components of ribosomes and the structures of membranes. In general, the relatedness of groups of organisms is reflected in the similarity of their DNA sequences—exactly the pattern that would be expected from descent and diversification from a common ancestor.

DNA sequences have also shed light on some of the mechanisms of evolution. For example, it is clear that the evolution of new functions for proteins commonly occurs after gene duplication events. These duplications are a kind of mutation in which an entire gene is added as an extra copy (or many copies) in the genome. These duplications allow the free modification of one copy by mutation, selection, and drift, while the second copy continues to produce a functional protein. This allows the original function

for the protein to be kept, while evolutionary forces tweak the copy until it functions in a new way.

Section Summary

The evidence for evolution is found at all levels of organization in living things and in the extinct species we know about through fossils. Fossils provide evidence for the evolutionary change through now extinct forms that led to modern species. For example, there is a rich fossil record that shows the evolutionary transitions from horse ancestors to modern horses that document intermediate forms and a gradual adaptation o changing ecosystems. The anatomy of species and the embryological development of that anatomy reveal common structures in divergent lineages that have been modified over time by evolution. The geographical distribution of living species reflects the origins of species in particular geographic locations and the history of continental movements. The structures of molecules, like anatomical structures, reflect the relationships of living species and match patterns of similarity expected from descent with modification.

Multiple Choice

Exercise:

Problem:

The wing of a bird and the arm of a human are examples of _________.

a. vestigial structures
b. molecular structures
c. homologous structures
d. analogous structures

Solution:

C

Exercise:

Problem:

The fact that DNA sequences are more similar in more closely related organisms is evidence of what?

a. optimal design in organisms
b. adaptation
c. mutation
d. descent with modification

Solution:

D

Free Response

Exercise:

Problem:

Why do scientists consider vestigial structures evidence for evolution?

Solution:

A vestigial structure is an example of a homologous structure that has apparently been reduced through evolution to a non-functional state because its function is no longer utilized by the species exhibiting it; therefore, any mutations which might reduce its structure are not selected against. The fact that the species has vestiges of the structure rather than no structure at all is evidence that it was present in an ancestor and evolved to non-functionality through accumulation of random mutations.

Glossary

vestigial structure

a physical structure present in an organism but that has no apparent function and appears to be from a functional structure in a distant ancestor

Speciation

By the end of this section, you will be able to:

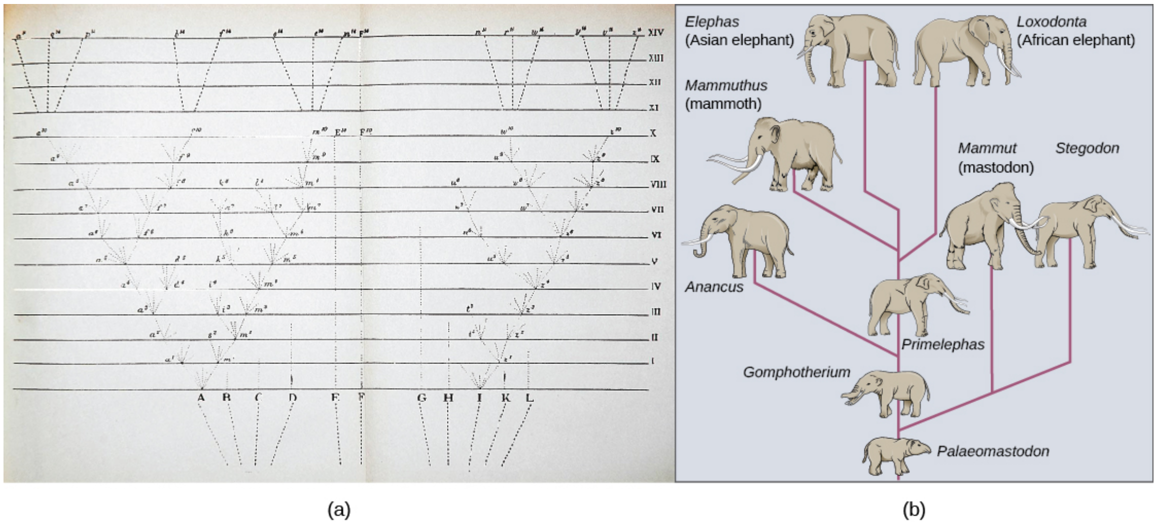
- Describe the definition of species and how species are identified as different
- Explain allopatric and sympatric speciation
- Describe adaptive radiation

The biological definition of species, which works for sexually reproducing organisms, is a group of actually or potentially interbreeding individuals. According to this definition, one species is distinguished from another by the possibility of matings between individuals from each species to produce fertile offspring. There are exceptions to this rule. Many species are similar enough that hybrid offspring are possible and may often occur in nature, but for the majority of species this rule generally holds. In fact, the presence of hybrids between similar species suggests that they may have descended from a single interbreeding species and that the speciation process may not yet be completed.

Given the extraordinary diversity of life on the planet there must be mechanisms for **speciation**: the formation of two species from one original species. Darwin envisioned this process as a branching event and diagrammed the process in the only illustration found in *On the Origin of Species* ([link]**a**). For speciation to occur, two new populations must be formed from one original population, and they must evolve in such a way that it becomes impossible for individuals from the two new populations to interbreed. Biologists have proposed mechanisms by which this could occur that fall into two broad categories. **Allopatric speciation**, meaning speciation in "other homelands," involves a geographic separation of populations from a parent species and subsequent evolution. **Sympatric speciation**, meaning speciation in the "same homeland," involves speciation occurring within a parent species while remaining in one location.

Biologists think of speciation events as the splitting of one ancestral species into two descendant species. There is no reason why there might not be more than two species formed at one time except that it is less likely and

such multiple events can also be conceptualized as single splits occurring close in time.



The only illustration in Darwin’s *On the Origin of Species* is (a) a diagram showing speciation events leading to biological diversity. The diagram shows similarities to phylogenetic charts that are drawn today to illustrate the relationships of species. (b) Modern elephants evolved from the *Palaeomastodon*, a species that lived in Egypt 35–50 million years ago.

Speciation through Geographic Separation

A geographically continuous population has a gene pool that is relatively homogeneous. Gene flow, the movement of alleles across the range of the species, is relatively free because individuals can move and then mate with individuals in their new location. Thus, the frequency of an allele at one end of a distribution will be similar to the frequency of the allele at the other end. When populations become geographically discontinuous that free-flow of alleles is prevented. When that separation lasts for a period of time, the two populations are able to evolve along different trajectories. Thus, their allele frequencies at numerous genetic loci gradually become more and more different as new alleles independently arise by mutation in each population. Typically, environmental conditions, such as climate, resources,

predators, and competitors, for the two populations will differ causing natural selection to favor divergent adaptations in each group. Different histories of genetic drift, enhanced because the populations are smaller than the parent population, will also lead to divergence.

Given enough time, the genetic and phenotypic divergence between populations will likely affect characters that influence reproduction enough that were individuals of the two populations brought together, mating would be less likely, or if a mating occurred, offspring would be non-viable or infertile. Many types of diverging characters may affect the reproductive isolation (inability to interbreed) of the two populations. These mechanisms of reproductive isolation can be divided into prezygotic mechanisms (those that operate before fertilization) and postzygotic mechanisms (those that operate after fertilization). Prezygotic mechanisms include traits that allow the individuals to find each other, such as the timing of mating, sensitivity to pheromones, or choice of mating sites. If individuals are able to encounter each other, character divergence may prevent courtship rituals from leading to a mating either because female preferences have changed or male behaviors have changed. Physiological changes may interfere with successful fertilization if mating is able to occur. Postzygotic mechanisms include genetic incompatibilities that prevent proper development of the offspring, or if the offspring live, they may be unable to produce viable gametes themselves as in the example of the mule, the infertile offspring of a female horse and a male donkey.

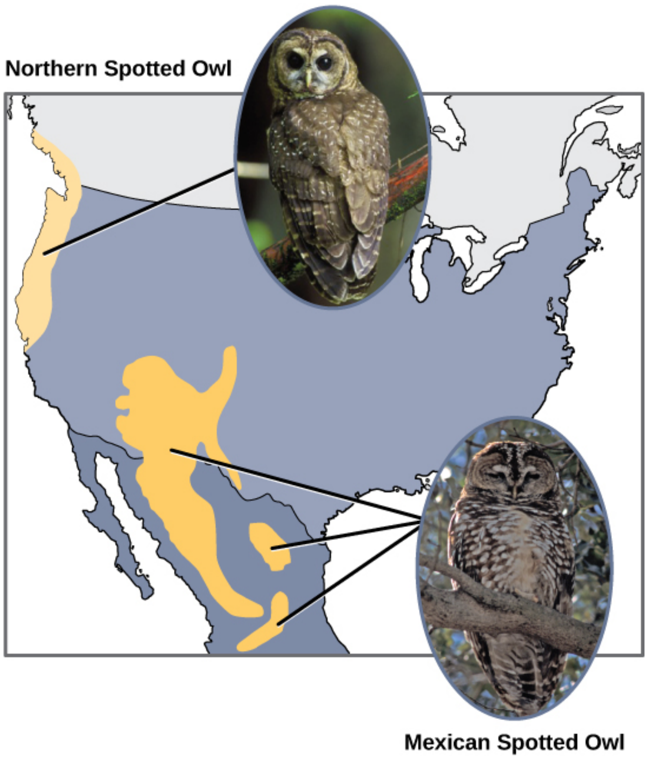
If the two isolated populations are brought back together and the hybrid offspring that formed from matings between individuals of the two populations have lower survivorship or reduced fertility, then selection will favor individuals that are able to discriminate between potential mates of their own population and the other population. This selection will enhance the reproductive isolation.

Isolation of populations leading to allopatric speciation can occur in a variety of ways: from a river forming a new branch, erosion forming a new valley, or a group of organisms traveling to a new location without the ability to return, such as seeds floating over the ocean to an island. The nature of the geographic separation necessary to isolate populations

depends entirely on the biology of the organism and its potential for dispersal. If two flying insect populations took up residence in separate nearby valleys, chances are that individuals from each population would fly back and forth, continuing gene flow. However, if two rodent populations became divided by the formation of a new lake, continued gene flow would be unlikely; therefore, speciation would be more likely.

Biologists group allopatric processes into two categories. If a few members of a species move to a new geographical area, this is called **dispersal**. If a natural situation arises to physically divide organisms, this is called **vicariance**.

Scientists have documented numerous cases of allopatric speciation taking place. For example, along the west coast of the United States, two separate subspecies of spotted owls exist. The northern spotted owl has genetic and phenotypic differences from its close relative, the Mexican spotted owl, which lives in the south ([link]). The cause of their initial separation is not clear, but it may have been caused by the glaciers of the ice age dividing an initial population into two.[footnote]
Courtney, S.P., et al, “Scientific Evaluation of the Status of the Northern Spotted Owl,” Sustainable Ecosystems Institute (2004), Portland, OR.

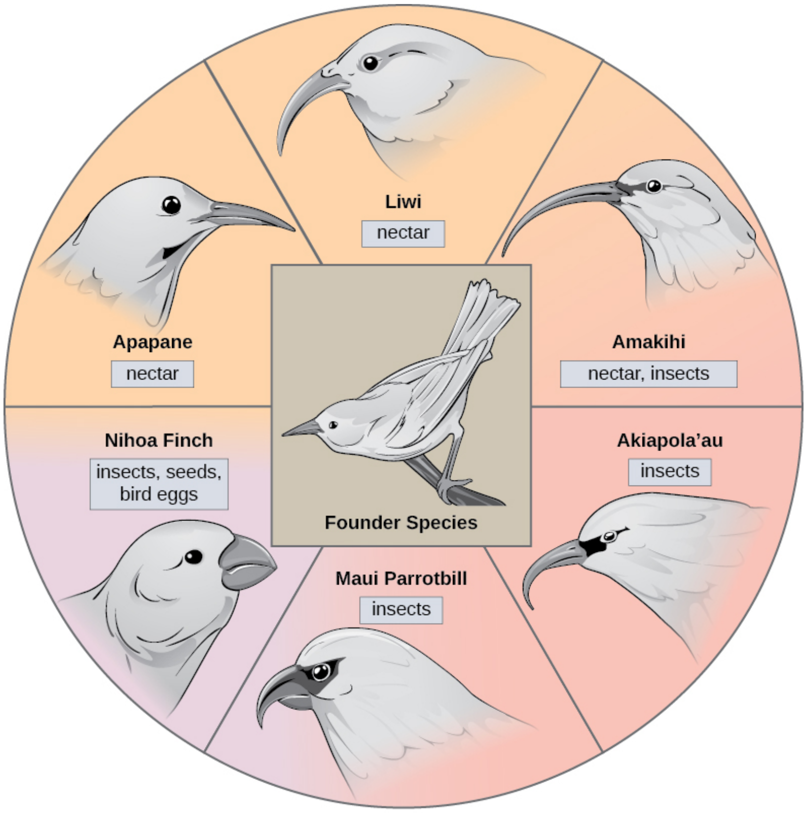


The northern spotted owl and the Mexican spotted owl inhabit geographically separate locations with different climates and ecosystems. The owl is an example of incipient speciation. (credit "northern spotted owl": modification of work by John and Karen Hollingsworth, USFWS; credit "Mexican spotted owl": modification of work by Bill Radke, USFWS)

Additionally, scientists have found that the further the distance between two groups that once were the same species, the more likely for speciation to take place. This seems logical because as the distance increases, the various environmental factors would likely have less in common than locations in close proximity. Consider the two owls; in the north, the climate is cooler

than in the south; the other types of organisms in each ecosystem differ, as do their behaviors and habits; also, the hunting habits and prey choices of the owls in the south vary from the northern ones. These variances can lead to evolved differences in the owls, and over time speciation will likely occur unless gene flow between the populations is restored.

In some cases, a population of one species disperses throughout an area, and each finds a distinct niche or isolated habitat. Over time, the varied demands of their new lifestyles lead to multiple speciation events originating from a single species, which is called **adaptive radiation**. From one point of origin, many adaptations evolve causing the species to radiate into several new ones. Island archipelagos like the Hawaiian Islands provide an ideal context for adaptive radiation events because water surrounds each island, which leads to geographical isolation for many organisms ([link]). The Hawaiian honeycreeper illustrates one example of adaptive radiation. From a single species, called the founder species, numerous species have evolved, including the eight shown in [link].



The honeycreeper birds illustrate adaptive radiation. From one original species of bird, multiple others evolved, each with its own distinctive characteristics.

Notice the differences in the species' beaks in [link]. Change in the genetic variation for beaks in response to natural selection based on specific food sources in each new habitat led to evolution of a different beak suited to the specific food source. The fruit and seed-eating birds have thicker, stronger beaks which are suited to break hard nuts. The nectar-eating birds have long beaks to dip into flowers to reach their nectar. The insect-eating birds have beaks like swords, appropriate for stabbing and impaling insects. Darwin's finches are another well-studied example of adaptive radiation in an archipelago.

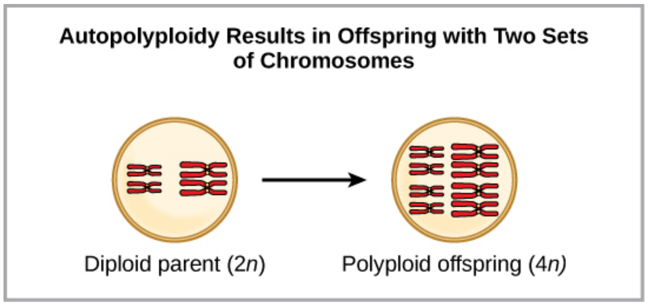
Note:
Concept in Action

Click through this interactive site to see how island birds evolved; click to see images of each species in evolutionary increments from five million years ago to today.

Speciation without Geographic Separation

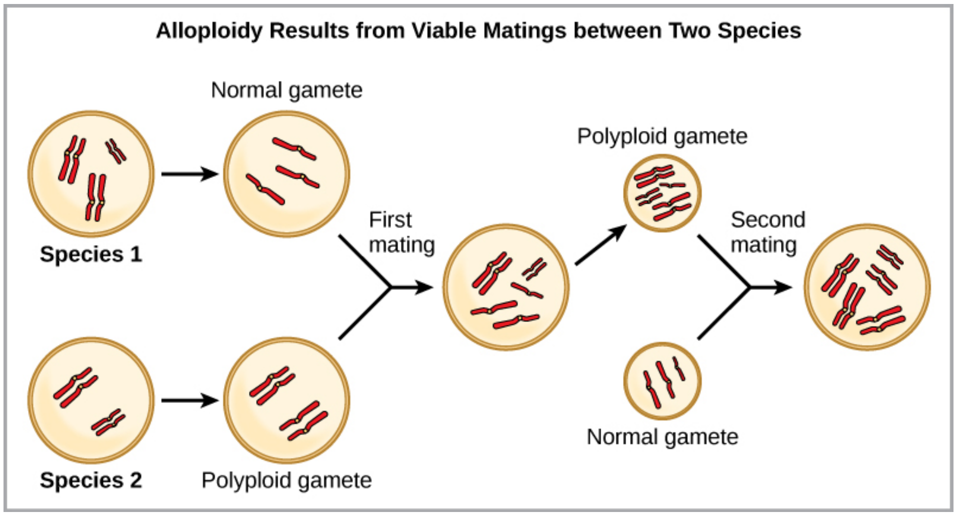
Can divergence occur if no physical barriers are in place to separate individuals who continue to live and reproduce in the same habitat? A number of mechanisms for sympatric speciation have been proposed and studied.

One form of sympatric speciation can begin with a chromosomal error during meiosis or the formation of a hybrid individual with too many chromosomes. Polyploidy is a condition in which a cell, or organism, has an extra set, or sets, of chromosomes. Scientists have identified two main types of polyploidy that can lead to reproductive isolation of an individual in the polyploid state. In some cases a polyploid individual will have two or more complete sets of chromosomes from its own species in a condition called autopolyploidy ([link]). The prefix “auto” means self, so the term means multiple chromosomes from one’s own species. Polyploidy results from an error in meiosis in which all of the chromosomes move into one cell instead of separating.



Autopolyploidy results when mitosis is not followed by cytokinesis.

For example, if a plant species with $2n = 6$ produces autopolyploid gametes that are also diploid ($2n = 6$, when they should be $n = 3$), the gametes now have twice as many chromosomes as they should have. These new gametes will be incompatible with the normal gametes produced by this plant species. But they could either self-pollinate or reproduce with other autopolyploid plants with gametes having the same diploid number. In this way, sympatric speciation can occur quickly by forming offspring with $4n$ called a tetraploid. These individuals would immediately be able to reproduce only with those of this new kind and not those of the ancestral species. The other form of polyploidy occurs when individuals of two different species reproduce to form a viable offspring called an allopolyploid. The prefix “allo” means “other” (recall from allopatric); therefore, an allopolyploid occurs when gametes from two different species combine. [link] illustrates one possible way an allopolyploidy can form. Notice how it takes two generations, or two reproductive acts, before the viable fertile hybrid results.



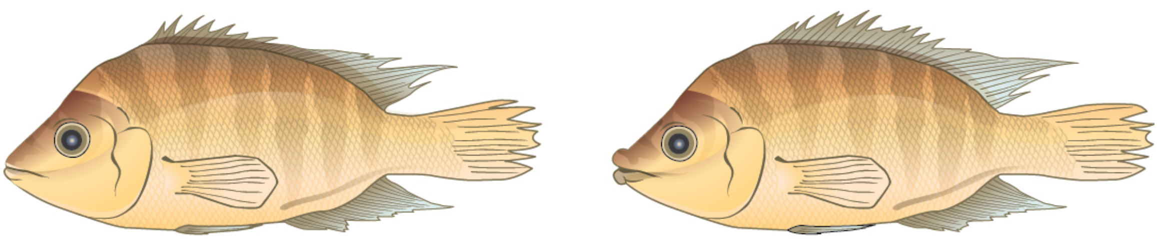
Alloploidy results when two species mate to produce viable offspring. In the example shown, a normal gamete from one species fuses with a polyploid gamete from another. Two matings are necessary to produce viable offspring.

The cultivated forms of wheat, cotton, and tobacco plants are all allopolyploids. Although polyploidy occurs occasionally in animals, most chromosomal abnormalities in animals are lethal; it takes place most commonly in plants. Scientists have discovered more than 1/2 of all plant species studied relate back to a species evolved through polyploidy.

Sympatric speciation may also take place in ways other than polyploidy. For example, imagine a species of fish that lived in a lake. As the population grew, competition for food also grew. Under pressure to find food, suppose that a group of these fish had the genetic flexibility to discover and feed off another resource that was unused by the other fish. What if this new food source was found at a different depth of the lake? Over time, those feeding on the second food source would interact more with each other than the other fish; therefore they would breed together as well. Offspring of these fish would likely behave as their parents and feed and live in the same area, keeping them separate from the original population. If this group of fish continued to remain separate from the first

population, eventually sympatric speciation might occur as more genetic differences accumulated between them.

This scenario does play out in nature, as do others that lead to reproductive isolation. One such place is Lake Victoria in Africa, famous for its sympatric speciation of cichlid fish. Researchers have found hundreds of sympatric speciation events in these fish, which have not only happened in great number, but also over a short period of time. [link] shows this type of speciation among a cichlid fish population in Nicaragua. In this locale, two types of cichlids live in the same geographic location; however, they have come to have different morphologies that allow them to eat various food sources.



Cichlid fish from Lake Apoyeque, Nicaragua, show evidence of sympatric speciation. Lake Apoyeque, a crater lake, is 1800 years old, but genetic evidence indicates that the lake was populated only 100 years ago by a single population of cichlid fish. Nevertheless, two populations with distinct morphologies and diets now exist in the lake, and scientists believe these populations may be in an early stage of speciation.

Finally, a well-documented example of ongoing sympatric speciation occurred in the apple maggot fly, *Rhagoletis pomonella*, which arose as an isolated population sometime after the introduction of the apple into North America. The native population of flies fed on hawthorn species and is host-specific: it only infests hawthorn trees. Importantly, it also uses the trees as a location to meet for mating. It is hypothesized that either through

mutation or a behavioral mistake, flies jumped hosts and met and mated in apple trees, subsequently laying their eggs in apple fruit. The offspring matured and kept their preference for the apple trees effectively dividing the original population into two new populations separated by host species, not by geography. The host jump took place in the nineteenth century, but there are now measureable differences between the two populations of fly. It seems likely that host specificity of parasites in general is a common cause of sympatric speciation.

Section Summary

Speciation occurs along two main pathways: geographic separation (allopatric speciation) and through mechanisms that occur within a shared habitat (sympatric speciation). Both pathways force reproductive isolation between populations. Sympatric speciation can occur through errors in meiosis that form gametes with extra chromosomes, called polyploidy. Autopolyploidy occurs within a single species, whereas allopolyploidy occurs because of a mating between closely related species. Once the populations are isolated, evolutionary divergence can take place leading to the evolution of reproductive isolating traits that prevent interbreeding should the two populations come together again. The reduced viability of hybrid offspring after a period of isolation is expected to select for stronger inherent isolating mechanisms.

Multiple Choice

Exercise:

Problem:

Which situation would most likely lead to allopatric speciation?

a. A flood causes the formation of a new lake.
b. A storm causes several large trees to fall down.
c. A mutation causes a new trait to develop.
d. An injury causes an organism to seek out a new food source.

Solution:

A

Exercise:

Problem:

What is the main difference between dispersal and vicariance?

a. One leads to allopatric speciation, whereas the other leads to sympatric speciation.
b. One involves the movement of the organism, whereas the other involves a change in the environment.
c. One depends on a genetic mutation occurring, whereas the other does not.
d. One involves closely related organisms, whereas the other involves only individuals of the same species.

Solution:

B

Exercise:

Problem:

Which variable increases the likelihood of allopatric speciation taking place more quickly?

a. lower rate of mutation
b. longer distance between divided groups
c. increased instances of hybrid formation
d. equivalent numbers of individuals in each population

Solution:

B

Free Response

Exercise:

Problem:

Why do island chains provide ideal conditions for adaptive radiation to occur?

Solution:

Organisms of one species can arrive to an island together and then disperse throughout the chain, each settling into different niches, exploiting different food resources and, evolving independently with little gene flow between different islands.

Exercise:

Problem:

Two species of fish had recently undergone sympatric speciation. The males of each species had a different coloring through which females could identify and choose a partner from her own species. After some time, pollution made the lake so cloudy it was hard for females to distinguish colors. What might take place in this situation?

Solution:

It is likely the two species would start to reproduce with each other if hybridization is still possible. Depending on the viability of their offspring, they may fuse back into one species.

Glossary

adaptive radiation
: a speciation when one species radiates out to form several other species

allopatric speciation

a speciation that occurs via a geographic separation

dispersal
: an allopatric speciation that occurs when a few members of a species move to a new geographical area

speciation
: a formation of a new species

sympatric speciation
: a speciation that occurs in the same geographic space

vicariance
: an allopatric speciation that occurs when something in the environment separates organisms of the same species into separate groups

Common Misconceptions about Evolution
By the end of this section, you will be able to:

- Identify common misconceptions about evolution
- Identify common criticisms of evolution

Although the theory of evolution initially generated some controversy, by 20 years after the publication of *On the Origin of Species* it was almost universally accepted by biologists, particularly younger biologists. Nevertheless, the theory of evolution is a difficult concept and misconceptions about how it works abound. In addition, there are those that reject it as an explanation for the diversity of life.

Note:
Concept in Action

This website addresses some of the main misconceptions associated with the theory of evolution.

Evolution Is Just a Theory

Critics of the theory of evolution dismiss its importance by purposefully confounding the everyday usage of the word “theory” with the way scientists use the word. In science, a “theory” is understood to be a concept that has been extensively tested and supported over time. We have a theory of the atom, a theory of gravity, and the theory of relativity, each of which describes what scientists understand to be facts about the world. In the same way, the theory of evolution describes facts about the living world. As such,

a theory in science has survived significant efforts to discredit it by scientists, who are naturally skeptical. While theories can sometimes be overturned or revised, this does not lessen their weight but simply reflects the constantly evolving state of scientific knowledge. In contrast, a "theory" in common vernacular means a guess or suggested explanation for something. This meaning is more akin to the concept of a "hypothesis" used by scientists, which is a tentative explanation for something that is proposed to either be supported or disproved. When critics of evolution say evolution is "just a theory," they are implying that there is little evidence supporting it and that it is still in the process of being rigorously tested. This is a mischaracterization. If this were the case, geneticist Theodosius Dobzhansky would not have said that "nothing in biology makes sense, except in the light of evolution."[footnote]

Theodosius Dobzhansky. "Biology, Molecular and Organismic." *American Zoologist* 4, no. 4 (1964): 449.

Individuals Evolve

An individual is born with the genes it has—these do not change as the individual ages. Therefore, an individual cannot evolve or adapt through natural selection. Evolution is the change in genetic composition of a population over time, specifically over generations, resulting from differential reproduction of individuals with certain alleles. Individuals do change over their lifetime, but this is called development; it involves changes programmed by the set of genes the individual acquired at birth in coordination with the individual's environment. When thinking about the evolution of a characteristic, it is probably best to think about the change of the average value of the characteristic in the population over time. For example, when natural selection leads to bill-size change in medium ground finches in the Galápagos, this does not mean that individual bills on the finches are changing. If one measures the average bill size among all individuals in the population at one time, and then measures the average bill size in the population several years later after there has been a strong selective pressure, this average value may be different as a result of evolution. Although some individuals may survive from the first time to the second, those individuals will still have the same bill size. However, there

may be enough new individuals with different bill sizes to change the average bill size.

Evolution Explains the Origin of Life

It is a common misunderstanding that evolution includes an explanation of life's origins. Conversely, some of the theory's critics complain that it cannot explain the origin of life. The theory does not try to explain the origin of life. The theory of evolution explains how populations change over time and how life diversifies—the origin of species. It does not shed light on the beginnings of life including the origins of the first cells, which is how life is defined. The mechanisms of the origin of life on Earth are a particularly difficult problem because it occurred a very long time ago, over a very long time, and presumably just occurred once. Importantly, biologists believe that the presence of life on Earth precludes the possibility that the events that led to life on Earth can be repeated because the intermediate stages would immediately become food for existing living things. The early stages of life included the formation of organic molecules such as carbohydrates, amino acids, or nucleotides. If these were formed from inorganic precursors today, they would simply be broken down by living things. The early stages of life also probably included more complex aggregations of molecules into enclosed structures with an internal environment, a boundary layer of some form, and the external environment. Such structures, if they were formed now, would be quickly consumed or broken down by living organisms.

However, once a mechanism of inheritance was in place in the form of a molecule like DNA or RNA, either within a cell or within a pre-cell, these entities would be subject to the principle of natural selection. More effective reproducers would increase in frequency at the expense of inefficient reproducers. So while evolution does not explain the origin of life, it may have something to say about some of the processes operating once pre-living entities acquired certain properties.

Organisms Evolve on Purpose

Statements such as “organisms evolve in response to a change in an environment,” are quite common. There are two easy misunderstandings possible with such a statement. First of all, the statement must not be understood to mean that individual organisms evolve, as was discussed above. The statement is shorthand for “a population evolves in response to a changing environment.” However, a second misunderstanding may arise by interpreting the statement to mean that the evolution is somehow intentional. A changed environment results in some individuals in the population, those with particular phenotypes, benefiting and, therefore, producing proportionately more offspring than other phenotypes. This results in change in the population if the characters are genetically determined.

It is also important to understand that the variation that natural selection works on is already in a population and does not arise in response to an environmental change. For example, applying antibiotics to a population of bacteria will, over time, select for a population of bacteria that are resistant to antibiotics. The resistance, which is caused by a gene, did not arise by mutation because of the application of the antibiotic. The gene for resistance was already present in the gene pool of the bacteria, likely at a low frequency. The antibiotic, which kills the bacterial cells without the resistance gene, strongly selects for individuals that are resistant, since these would be the only ones that survived and divided. Experiments have demonstrated that mutations for antibiotic resistance do not arise as a result of antibiotic application.

In a larger sense, evolution is also not goal directed. Species do not become “better” over time; they simply track their changing environment with adaptations that maximize their reproduction in a particular environment at a particular time. Evolution has no goal of making faster, bigger, more complex, or even smarter species. This kind of language is common in popular literature. Certain organisms, ourselves included, are described as the “pinnacle” of evolution, or “perfected” by evolution. What characteristics evolve in a species are a function of the variation present and the environment, both of which are constantly changing in a non-directional way. What trait is fit in one environment at one time may well be fatal at

some point in the future. This holds equally well for a species of insect as it does the human species.

Evolution Is Controversial among Scientists

The theory of evolution was controversial when it was first proposed in 1859, yet within 20 years virtually every working biologist had accepted evolution as the explanation for the diversity of life. The rate of acceptance was extraordinarily rapid, partly because Darwin had amassed an impressive body of evidence. The early controversies involved both scientific arguments against the theory and the arguments of religious leaders. It was the arguments of the biologists that were resolved after a short time, while the arguments of religious leaders have persisted to this day.

The theory of evolution replaced the predominant theory at the time that species had all been specially created within relatively recent history. Despite the prevalence of this theory, it was becoming increasingly clear to naturalists during the nineteenth century that it could no longer explain many observations of geology and the living world. The persuasiveness of the theory of evolution to these naturalists lay in its ability to explain these phenomena, and it continues to hold extraordinary explanatory power to this day. Its continued rejection by some religious leaders results from its replacement of special creation, a tenet of their religious belief. These leaders cannot accept the replacement of special creation by a mechanistic process that excludes the actions of a deity as an explanation for the diversity of life including the origins of the human species. It should be noted, however, that most of the major denominations in the United States have statements supporting the acceptance of evidence for evolution as compatible with their theologies.

The nature of the arguments against evolution by religious leaders has evolved over time. One current argument is that the theory is still controversial among biologists. This claim is simply not true. The number of working scientists who reject the theory of evolution, or question its validity and say so, is small. A Pew Research poll in 2009 found that 97 percent of the 2500 scientists polled believe species evolve.[footnote] The

support for the theory is reflected in signed statements from many scientific societies such as the American Association for the Advancement of Science, which includes working scientists as members. Many of the scientists that reject or question the theory of evolution are non-biologists, such as engineers, physicians, and chemists. There are no experimental results or research programs that contradict the theory. There are no papers published in peer-reviewed scientific journals that appear to refute the theory. The latter observation might be considered a consequence of suppression of dissent, but it must be remembered that scientists are skeptics and that there is a long history of published reports that challenged scientific orthodoxy in unpopular ways. Examples include the endosymbiotic theory of eukaryotic origins, the theory of group selection, the microbial cause of stomach ulcers, the asteroid-impact theory of the Cretaceous extinction, and the theory of plate tectonics. Research with evidence and ideas with scientific merit are considered by the scientific community. Research that does not meet these standards is rejected.
Pew Research Center for the People & the Press, *Public Praises Science; Scientists Fault Public, Media* (Washington, DC, 2009), 37.

Other Theories Should Be Taught

A common argument from some religious leaders is that alternative theories to evolution should be taught in public schools. Critics of evolution use this strategy to create uncertainty about the validity of the theory without offering actual evidence. In fact, there are no viable alternative scientific theories to evolution. The last such theory, proposed by Lamarck in the nineteenth century, was replaced by the theory of natural selection. A single exception was a research program in the Soviet Union based on Lamarck’s theory during the early twentieth century that set that country’s agricultural research back decades. Special creation is not a viable alternative scientific theory because it is not a scientific theory, since it relies on an untestable explanation. Intelligent design, despite the claims of its proponents, is also not a scientific explanation. This is because intelligent design posits the existence of an unknown designer of living organisms and their systems. Whether the designer is unknown or supernatural, it is a cause that cannot be measured; therefore, it is not a scientific explanation. There are two reasons not to teach nonscientific theories. First, these explanations for the

diversity of life lack scientific usefulness because they do not, and cannot, give rise to research programs that promote our understanding of the natural world. Experiments cannot test non-material explanations for natural phenomena. For this reason, teaching these explanations as science in public schools is not in the public interest. Second, in the United States, it is illegal to teach them as science because the U.S. Supreme Court and lower courts have ruled that the teaching of religious belief, such as special creation or intelligent design, violates the establishment clause of the First Amendment of the U.S. Constitution, which prohibits government sponsorship of a particular religion.

The theory of evolution and science in general is, by definition, silent on the existence or non-existence of the spiritual world. Science is only able to study and know the material world. Individual biologists have sometimes been vocal atheists, but it is equally true that there are many deeply religious biologists. Nothing in biology precludes the existence of a god, indeed biology as a science has nothing to say about it. The individual biologist is free to reconcile her or his personal and scientific knowledge as they see fit. The Voices for Evolution project (http://ncse.com/voices), developed through the National Center for Science Education, works to gather the diversity of perspectives on evolution to advocate it being taught in public schools.

Section Summary

The theory of evolution is a difficult concept and misconceptions abound. The factual nature of evolution is often challenged by wrongly associating the scientific meaning of a theory with the vernacular meaning. Evolution is sometimes mistakenly interpreted to mean that individuals evolve, when in fact only populations can evolve as their gene frequencies change over time. Evolution is often assumed to explain the origin of life, which it does not speak to. It is often spoken in goal-directed terms by which organisms change through intention, and selection operates on mutations present in a population that have not arisen in response to a particular environmental stress. Evolution is often characterized as being controversial among scientists; however, it is accepted by the vast majority of working scientists. Critics of evolution often argue that alternative theories to evolution should

be taught in public schools; however, there are no viable alternative scientific theories to evolution. The alternative religious beliefs should not be taught as science because it cannot be proven, and in the United States it is unconstitutional. Science is silent on the question of the existence of a god while scientists are able to reconcile religious belief and scientific knowledge.

Multiple Choice

Exercise:

Problem:

The word “theory” in theory of evolution is best replaced by ________.

a. fact
b. hypothesis
c. idea
d. alternate explanation

Solution:

A

Exercise:

Problem:

Why are alternative scientific theories to evolution not taught in public school?

a. more theories would confuse students
b. there are no viable scientific alternatives
c. it is against the law
d. alternative scientific theories are suppressed by the science establishment

---

Solution:

B

Free Response

Exercise:

Problem:

How does the scientific meaning of “theory” differ from the common, everyday meaning of the word?

Solution:

In science, a theory is a thoroughly tested and verified set of explanations for a body of observations of nature. It is the strongest form of knowledge in science. In contrast, a theory in common usage can mean a guess or speculation about something, meaning that the knowledge implied by the theory may be very weak.

Exercise:

Problem:

Explain why the statement that a monkey is more evolved than a mouse is incorrect.

Solution:

The statement implies that there is a goal to evolution and that the monkey represents greater progress to that goal than the mouse. Both species are likely to be well adapted to their particular environment, which is the outcome of natural selection.

Introduction
class="introduction"

Although they look different, this bee and flower are distantly related. (credit: modification of work by John Beetham)

This bee and *Echinacea* flower could not look more different, yet they are related, as are all living organisms on Earth. By following pathways of similarities and differences—both visible and genetic—scientists seek to map the history of evolution from single-celled organisms to the

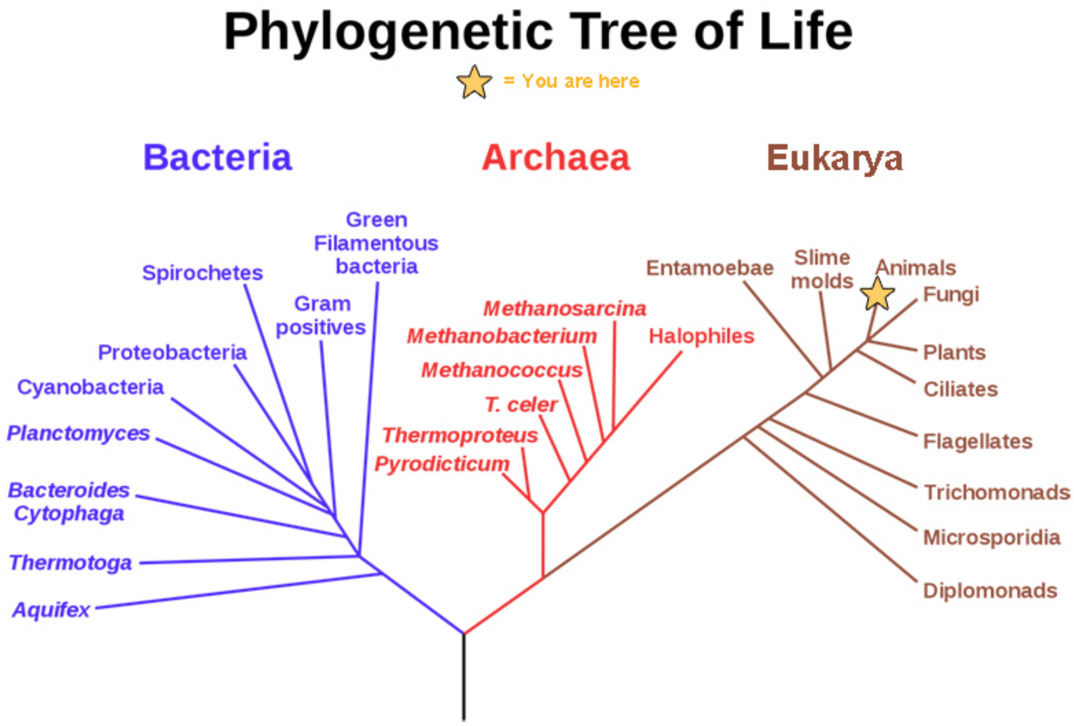
tremendous diversity of creatures that have crawled, germinated, floated, swam, flown, and walked on this planet.

Organizing Life on Earth

By the end of this section, you will be able to:

- Discuss the need for a comprehensive classification system
- List the different levels of the taxonomic classification system
- Describe how systematics and taxonomy relate to phylogeny

All life on Earth evolved from a common ancestor. Biologists map how organisms are related by constructing phylogenetic trees. In other words, a “tree of life” can be constructed to illustrate when different organisms evolved and to show the relationships among different organisms, as shown in [link]. Notice that from a single point, the three domains of Archaea, Bacteria, and Eukarya diverge and then branch repeatedly. The small branch that plants and animals (including humans) occupy in this diagram shows how recently these groups had their origin compared with other groups.



In the evolution of life on Earth, the three domains of life—Archaea, Bacteria, and Eukarya—branch from a single point. (credit: modification of work by Eric Gaba)

The phylogenetic tree in [link] illustrates the pathway of evolutionary history. The pathway can be traced from the origin of life to any individual species by navigating through the evolutionary branches between the two points. Also, by starting with a single species and tracing backward to any branch point, the organisms related to it by various degrees of closeness can be identified.

A **phylogeny** is the evolutionary history and the relationships among a species or group of species. The study of organisms with the purpose of deriving their relationships is called **systematics**.

Many disciplines within the study of biology contribute to understanding how past and present life evolved over time, and together they contribute to building, updating, and maintaining the "tree of life." Information gathered may include data collected from fossils, from studying morphology, from the structure of body parts, or from molecular structure, such as the sequence of amino acids in proteins or DNA nucleotides. By considering the trees generated by different sets of data scientists can put together the phylogeny of a species.

Scientists continue to discover new species of life on Earth as well as new character information, thus trees change as new data arrive.

The Levels of Classification

Taxonomy (which literally means "arrangement law") is the science of naming and grouping species to construct an internationally shared classification system. The taxonomic classification system (also called the Linnaean system after its inventor, Carl Linnaeus, a Swedish naturalist) uses a hierarchical model. A hierarchical system has levels and each group at one of the levels includes groups at the next lowest level, so that at the lowest level each member belongs to a series of nested groups. An analogy is the nested series of directories on the main disk drive of a computer. For example, in the most inclusive grouping, scientists divide organisms into three **domains**: Bacteria, Archaea, and Eukarya. Within each domain is a second level called a **kingdom**. Each domain contains several kingdoms.

Within kingdoms, the subsequent categories of increasing specificity are: **phylum**, **class**, **order**, **family**, **genus**, and **species**.

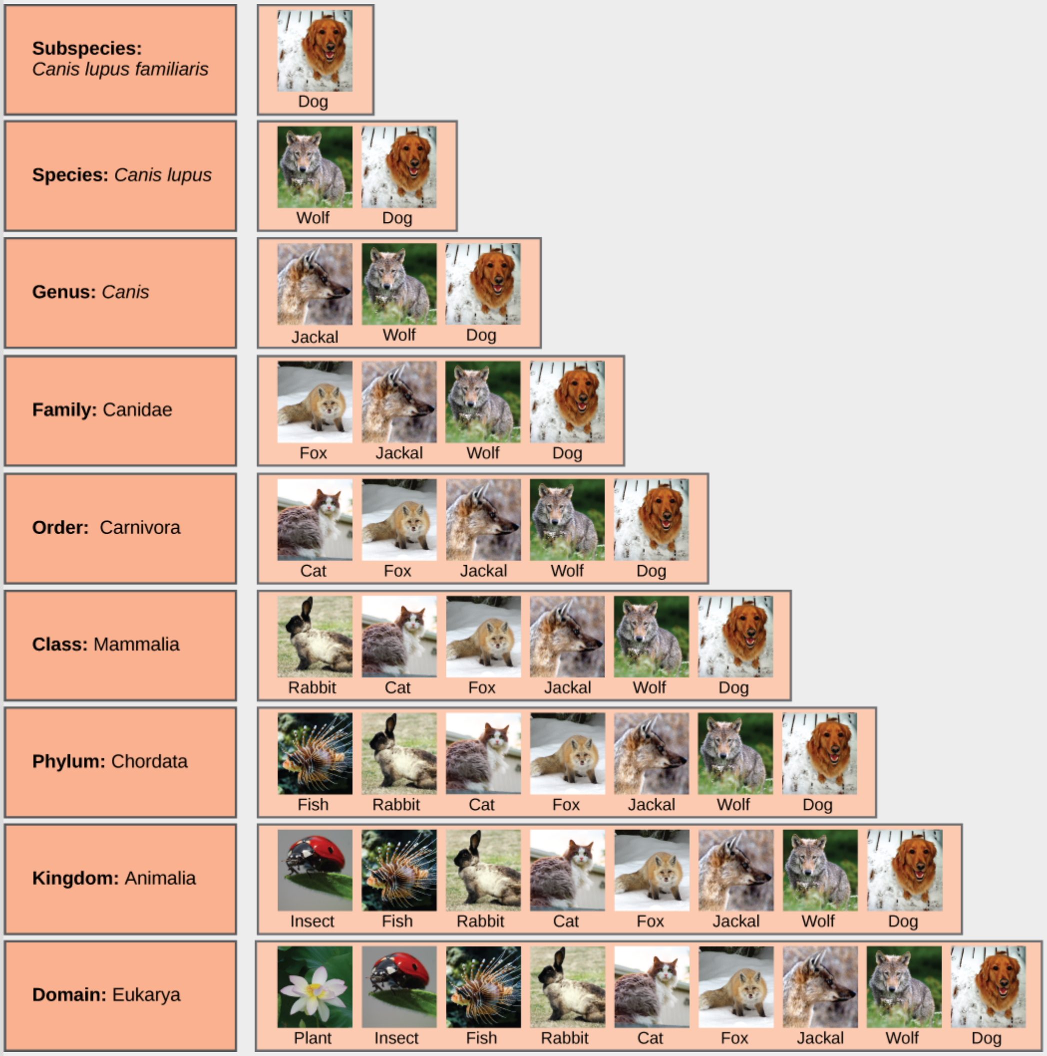
As an example, the classification levels for the domestic dog are shown in [link]. The group at each level is called a **taxon** (plural: taxa). In other words, for the dog, Carnivora is the taxon at the order level, Canidae is the taxon at the family level, and so forth. Organisms also have a common name that people typically use, such as domestic dog, or wolf. Each taxon name is capitalized except for species, and the genus and species names are italicized. Scientists refer to an organism by its genus and species names together, commonly called a scientific name, or Latin name. This two-name system is called **binomial nomenclature**. The scientific name of the wolf is therefore *Canis lupus*. Recent study of the DNA of domestic dogs and wolves suggest that the domestic dog is a subspecies of the wolf, not its own species, thus it is given an extra name to indicate its subspecies status, *Canis lupus familiaris*.

[link] also shows how taxonomic levels move toward specificity. Notice how within the domain we find the dog grouped with the widest diversity of organisms. These include plants and other organisms not pictured, such as fungi and protists. At each sublevel, the organisms become more similar because they are more closely related. Before Darwin's theory of evolution was developed, naturalists sometimes classified organisms using arbitrary similarities, but since the theory of evolution was proposed in the 19th century, biologists work to make the classification system reflect evolutionary relationships. This means that all of the members of a taxon should have a common ancestor and be more closely related to each other than to members of other taxa.

Recent genetic analysis and other advancements have found that some earlier taxonomic classifications do not reflect actual evolutionary relationships, and therefore, changes and updates must be made as new discoveries take place. One dramatic and recent example was the breaking apart of prokaryotic species, which until the 1970s were all classified as bacteria. Their division into Archaea and Bacteria came about after the recognition that their large genetic differences warranted their separation into two of three fundamental branches of life.

Note:

Art Connection



At each sublevel in the taxonomic classification system, organisms become more similar. Dogs and wolves are the same species because they can breed and produce viable offspring, but they are different enough to be classified as different subspecies. (credit “plant”: modification of work by "berduchwal"/Flickr; credit “insect”: modification of work by Jon Sullivan; credit “fish”: modification of work by Christian Mehlführer; credit “rabbit”: modification of work by Aidan Wojtas; credit “cat”: modification of work by Jonathan Lidbeck; credit “fox”: modification of work by Kevin Bacher, NPS; credit “jackal”: modification of work

by Thomas A. Hermann, NBII, USGS; credit “wolf” modification of work by Robert Dewar; credit “dog”: modification of work by "digital_image_fan"/Flickr)

In what levels are cats and dogs considered to be part of the same group?

Note:
Concept in Action

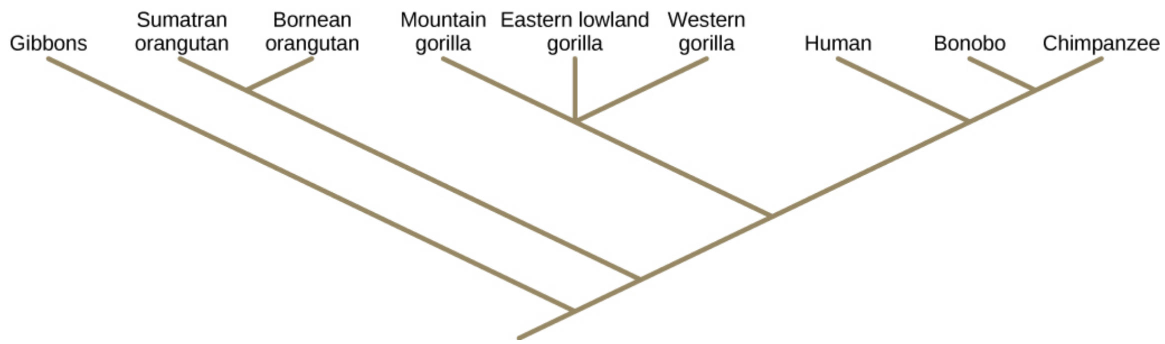
Visit this PBS site to learn more about taxonomy. Under Classifying Life, click Launch Interactive.

Classification and Phylogeny

Scientists use a tool called a phylogenetic tree to show the evolutionary pathways and relationships between organisms. A **phylogenetic tree** is a diagram used to reflect evolutionary relationships among organisms or groups of organisms. The hierarchical classification of groups nested within more inclusive groups is reflected in diagrams. Scientists consider phylogenetic trees to be a hypothesis of the evolutionary past because one cannot go back through time to confirm the proposed relationships.

Unlike with a taxonomic classification, a phylogenetic tree can be read like a map of evolutionary history, as shown in [link]. Shared characteristics are used to construct phylogenetic trees. The point where a split occurs in a tree, called a **branch point**, represents where a single lineage evolved into

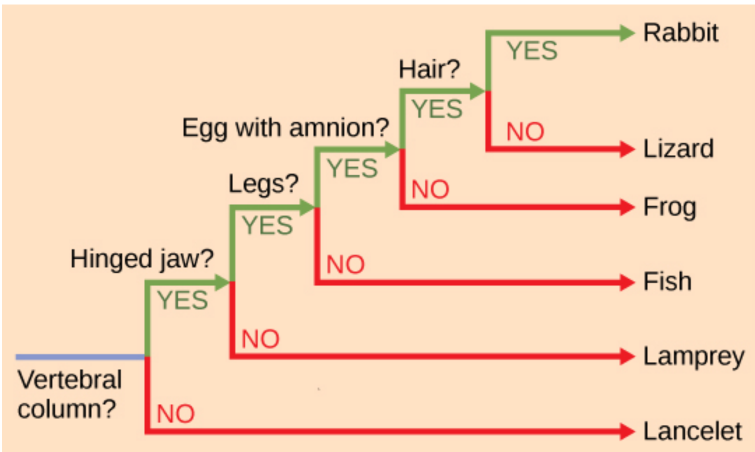
distinct new ones. Many phylogenetic trees have a single branch point at the base representing a common ancestor of all the branches in the tree. Scientists call such trees **rooted**, which means there is a single ancestral taxon at the base of a phylogenetic tree to which all organisms represented in the diagram descend from. When two lineages stem from the same branch point, they are called **sister taxa**, for example the two species of orangutans. A branch point with more than two groups illustrates a situation for which scientists have not definitively determined relationships. An example is illustrated by the three branches leading to the gorilla subspecies; their exact relationships are not yet understood. It is important to note that sister taxa share an ancestor, which does not mean that one taxon evolved from the other. The branch point, or split, represents a common ancestor that existed in the past, but that no longer exists. Humans did not evolve from chimpanzees (nor did chimpanzees evolve from humans) although they are our closest living relatives. Both humans and chimpanzees evolved from a common ancestor that lived, scientists believe, six million years ago and looked different from both modern chimpanzees and modern humans.



A phylogenetic tree is rooted and shows how different organisms, in this case the species and subspecies of living apes, evolved from a common ancestor.

The branch points and the branches in phylogenetic tree structure also imply evolutionary change. Sometimes the significant character changes are

identified on a branch or branch point. For example, in [link], the branch point that gives rise to the mammal and reptile lineage from the frog lineage shows the origin of the amniotic egg character. Also the branch point that gives rise to organisms with legs is indicated at the common ancestor of mammals, reptiles, amphibians, and jawed fishes.



This phylogenetic tree is rooted by an organism that lacked a vertebral column. At each branch point, organisms with different characters are placed in different groups.

Note:
Concept in Action

This interactive exercise allows you to explore the evolutionary relationships among species.

Limitations of Phylogenetic Trees

It is easy to assume that more closely related organisms look more alike, and while this is often the case, it is not always true. If two closely related lineages evolved under significantly different surroundings or after the evolution of a major new adaptation, they may look quite different from each other, even more so than other groups that are not as closely related. For example, the phylogenetic tree in [link] shows that lizards and rabbits both have amniotic eggs, whereas salamanders (within the frog lineage) do not; yet on the surface, lizards and salamanders appear more similar than the lizards and rabbits.

Another aspect of phylogenetic trees is that, unless otherwise indicated, the branches do not show length of time, they show only the order in time of evolutionary events. In other words, a long branch does not necessarily mean more time passed, nor does a short branch mean less time passed—unless specified on the diagram. For example, in [link], the tree does not indicate how much time passed between the evolution of amniotic eggs and hair. What the tree does show is the order in which things took place. Again using [link], the tree shows that the oldest trait is the vertebral column, followed by hinged jaws, and so forth. Remember that any phylogenetic tree is a part of the greater whole, and similar to a real tree, it does not grow in only one direction after a new branch develops. So, for the organisms in [link], just because a vertebral column evolved does not mean that invertebrate evolution ceased, it only means that a new branch formed. Also, groups that are not closely related, but evolve under similar conditions, may appear more similar to each other than to a close relative.

Section Summary

Scientists continually obtain new information that helps to understand the evolutionary history of life on Earth. Each group of organisms went through

its own evolutionary journey, called its phylogeny. Each organism shares relatedness with others, and based on morphologic and genetic evidence scientists attempt to map the evolutionary pathways of all life on Earth. Historically, organisms were organized into a taxonomic classification system. However, today many scientists build phylogenetic trees to illustrate evolutionary relationships and the taxonomic classification system is expected to reflect evolutionary relationships.

Art Connections

Exercise:

Problem:

[link] In what levels are cats and dogs considered to be part of the same group?

Solution:

[link] Cats and dogs are part of the same group at five levels: both are in the domain Eukarya, the kingdom Animalia, the phylum Chordata, the class Mammalia, and the order Carnivora.

Multiple Choice

Exercise:

Problem: What is a phylogeny a description of?

a. mutations
b. DNA
c. evolutionary history
d. organisms on Earth

Solution:

C

Exercise:

Problem: What do scientists in the field of systematics accomplish?

a. discover new fossil sites
b. organize and classify organisms
c. name new species
d. communicate between field biologists

Solution:

B

Exercise:

Problem:

Which statement about the taxonomic classification system is correct?

a. There are more domains than kingdoms.
b. Kingdoms are the top category of classification.
c. A phylum may be represented in more than one kingdom.
d. Species are the most specific category of classification.

Solution:

D

Exercise:

Problem:

Which best describes the relationship between chimpanzees and humans?

a. chimpanzees evolved from humans
b. humans evolved from chimpanzees

c. chimpanzees and humans evolved from a common ancestor
d. chimpanzees and humans belong to the same species

Solution:

C

Exercise:

Problem: Which best describes a branch point in a phylogenetic tree?

a. a hypothesis
b. new lineage
c. hybridization
d. a mating

Solution:

B

Free Response

Exercise:

Problem:

How does a phylogenetic tree indicate major evolutionary events within a lineage?

Solution:

The phylogenetic tree shows the order in which evolutionary events took place and in what order certain characteristics and organisms evolved in relation to others. It does not generally indicate time durations.

Exercise:

Problem:

List the different levels of the taxonomic classification system.

Solution:

Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species.

Glossary

binomial nomenclature
: a system of two-part scientific names for an organism, which includes genus and species names

branch point
: a point on a phylogenetic tree where a single lineage splits to distinct new ones

class
: the category in the taxonomic classification system that falls within phylum and includes orders

domain
: the highest level category in the classification system and that includes all taxonomic classifications below it; it is the most inclusive taxon

family
: the category in the taxonomic classification system that falls within order and includes genera

genus
: the category in the taxonomic classification system that falls within family and includes species; the first part of the scientific name

kingdom
: the category in the taxonomic classification system that falls within domain and includes phyla

order
: the category in the taxonomic classification system that falls within class and includes families

phylogenetic tree
: diagram used to reflect the evolutionary relationships between organisms or groups of organisms

phylogeny
: evolutionary history and relationship of an organism or group of organisms

phylum
: the category in the taxonomic classification system that falls within kingdom and includes classes

rooted
: describing a phylogenetic tree with a single ancestral lineage to which all organisms represented in the diagram relate

sister taxa
: two lineages that diverged from the same branch point

species
: the most specific category of classification

systematics
: the science of determining the evolutionary relationships of organisms

taxon
: a single level in the taxonomic classification system

taxonomy
: the science of classifying organisms

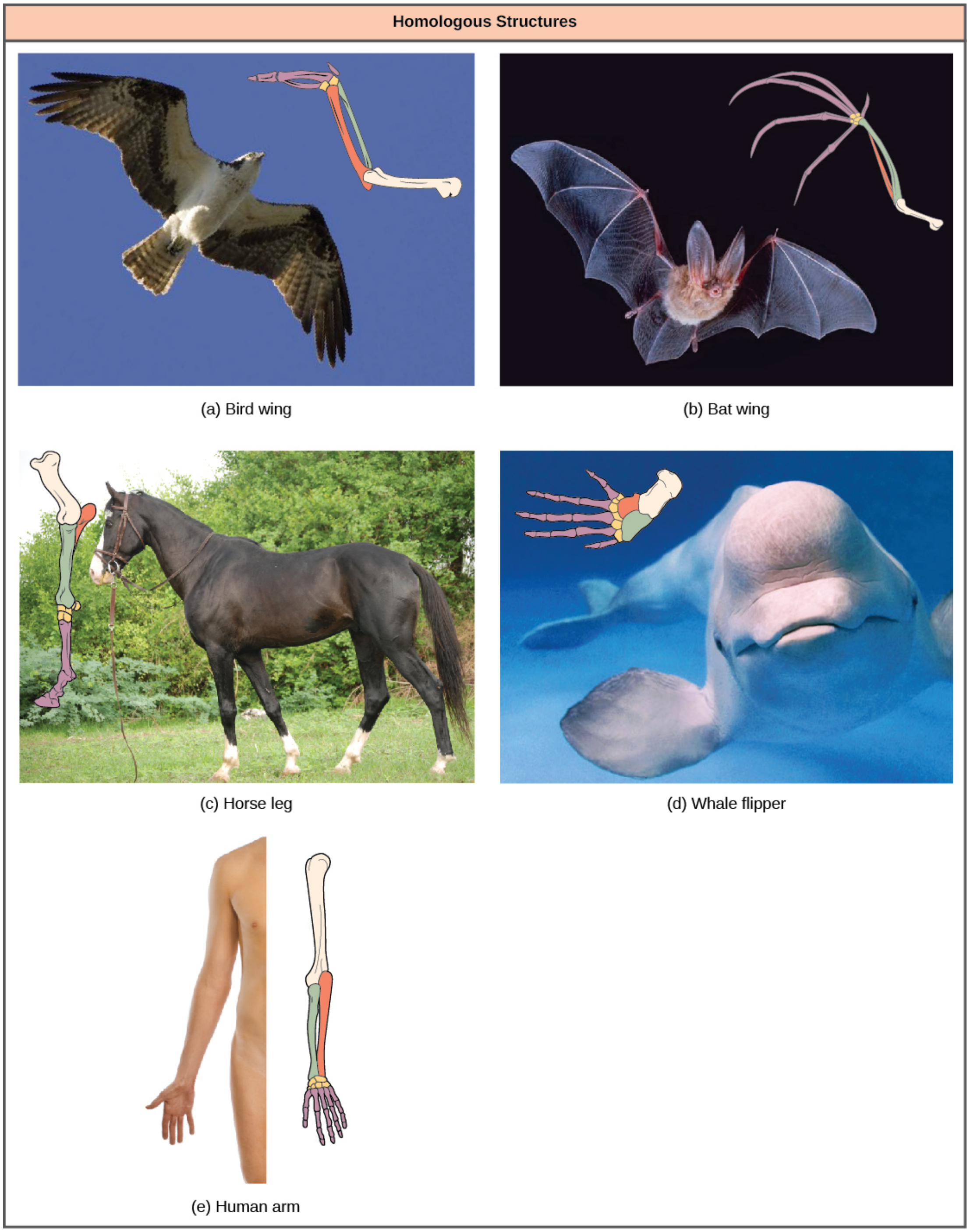
Determining Evolutionary Relationships
By the end of this section, you will be able to:

- Compare homologous and analogous traits
- Discuss the purpose of cladistics

Scientists collect information that allows them to make evolutionary connections between organisms. Similar to detective work, scientists must use evidence to uncover the facts. In the case of phylogeny, evolutionary investigations focus on two types of evidence: morphologic (form and function) and genetic.

Two Measures of Similarity

Organisms that share similar physical features *and* genetic sequences tend to be more closely related than those that do not. Features that overlap both morphologically and genetically are referred to as homologous structures; the similarities stem from common evolutionary paths. For example, as shown in [link], the bones in the wings of bats and birds, the arms of humans, and the foreleg of a horse are homologous structures. Notice the structure is not simply a single bone, but rather a grouping of several bones arranged in a similar way in each organism even though the elements of the structure may have changed shape and size.



Bat and bird wings, the foreleg of a horse, the flipper of a whale, and the arm of a human are homologous structures, indicating that bats, birds, horses, whales, and humans share a common evolutionary past. (credit a photo: modification of work by Steve Hillebrand, USFWS; credit b photo: modification of work by U.S. BLM; credit c photo:

modification of work by Virendra Kankariya; credit d photo: modification of work by Russian Gov./Wikimedia Commons)

Misleading Appearances

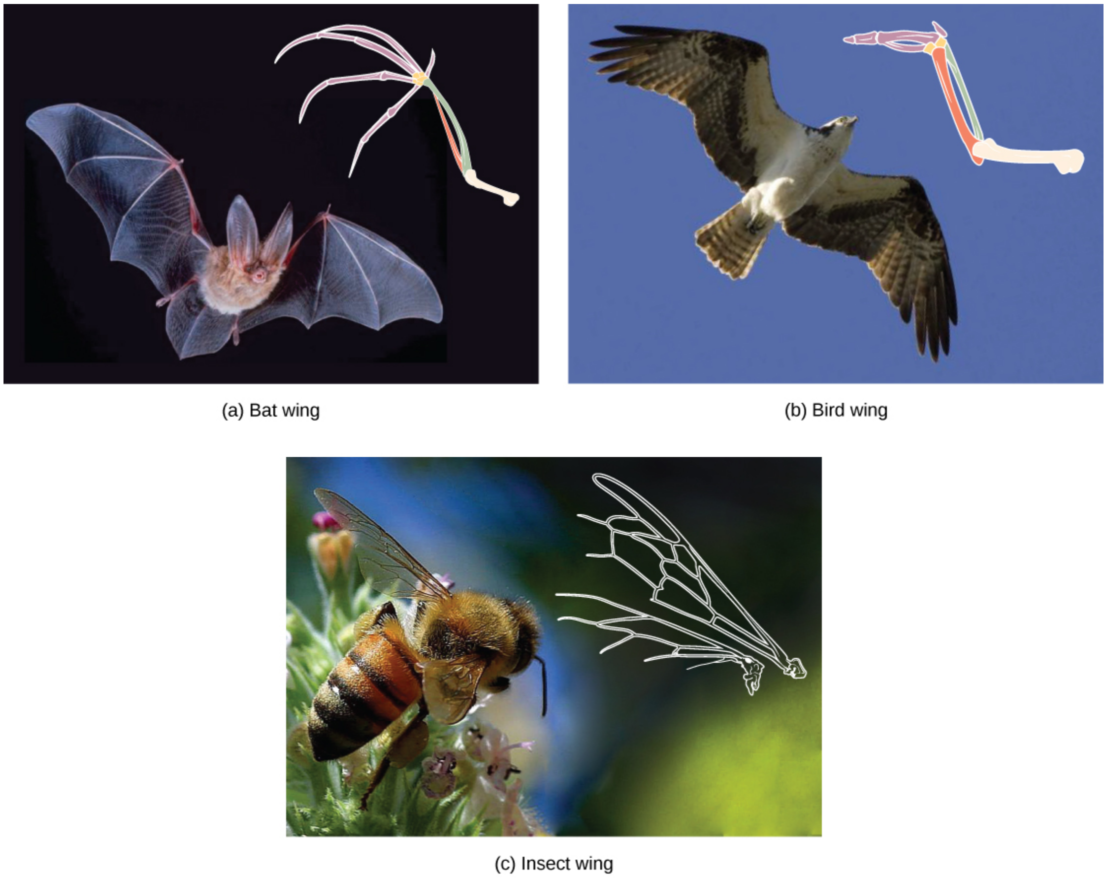
Some organisms may be very closely related, even though a minor genetic change caused a major morphological difference to make them look quite different. For example, chimpanzees and humans, the skulls of which are shown in [link] are very similar genetically, sharing 99 percent[footnote] of their genes. However, chimpanzees and humans show considerable anatomical differences, including the degree to which the jaw protrudes in the adult and the relative lengths of our arms and legs.
Gibbons, A. (2012, June 13). *Science Now*. Retrieved from http://news.sciencemag.org/sciencenow/2012/06/bonobo-genome-sequenced.html

(a) The chimpanzee jaw protrudes to a much greater

degree than (b) the human jaw. (credit a: modification of work by "Pastorius"/Wikimedia Commons)

However, unrelated organisms may be distantly related yet appear very much alike, usually because common adaptations to similar environmental conditions evolved in both. An example is the streamlined body shapes, the shapes of fins and appendages, and the shape of the tails in fishes and whales, which are mammals. These structures bear superficial similarity because they are adaptations to moving and maneuvering in the same environment—water. When a characteristic that is similar occurs by adaptive convergence (convergent evolution), and not because of a close evolutionary relationship, it is called an **analogous structure**. In another example, insects use wings to fly like bats and birds. We call them both wings because they perform the same function and have a superficially similar form, but the embryonic origin of the two wings is completely different. The difference in the development, or embryogenesis, of the wings in each case is a signal that insects and bats or birds do not share a common ancestor that had a wing. The wing structures, shown in [link] evolved independently in the two lineages.

Similar traits can be either homologous or analogous. Homologous traits share an evolutionary path that led to the development of that trait, and analogous traits do not. Scientists must determine which type of similarity a feature exhibits to decipher the phylogeny of the organisms being studied.



The wing of a honey bee is similar in shape to a bird wing and a bat wing and serves the same function (flight). The bird and bat wings are homologous structures. However, the honey bee wing has a different structure (it is made of a chitinous exoskeleton, not a boney endoskeleton) and embryonic origin. The bee and bird or bat wing types illustrate an analogy—similar structures that do not share an evolutionary history. (credit a photo: modification of work by U.S. BLM; credit b: modification of work by Steve Hillebrand, USFWS; credit c: modification of work by Jon Sullivan)

Note:
Concept in Action

This website has several examples to show how appearances can be misleading in understanding the phylogenetic relationships of organisms.

Molecular Comparisons

With the advancement of DNA technology, the area of **molecular systematics**, which describes the use of information on the molecular level including DNA sequencing, has blossomed. New analysis of molecular characters not only confirms many earlier classifications, but also uncovers previously made errors. Molecular characters can include differences in the amino-acid sequence of a protein, differences in the individual nucleotide sequence of a gene, or differences in the arrangements of genes. Phylogenies based on molecular characters assume that the more similar the sequences are in two organisms, the more closely related they are. Different genes change evolutionarily at different rates and this affects the level at which they are useful at identifying relationships. Rapidly evolving sequences are useful for determining the relationships among closely related species. More slowly evolving sequences are useful for determining the relationships between distantly related species. To determine the relationships between very different species such as Eukarya and Archaea, the genes used must be very ancient, slowly evolving genes that are present in both groups, such as the genes for ribosomal RNA. Comparing phylogenetic trees using different sequences and finding them similar helps to build confidence in the inferred relationships.

Sometimes two segments of DNA in distantly related organisms randomly share a high percentage of bases in the same locations, causing these organisms to appear closely related when they are not. For example, the fruit fly shares 60 percent of its DNA with humans.[footnote] In this situation, computer-based statistical algorithms have been developed to help identify

the actual relationships, and ultimately, the coupled use of both morphologic and molecular information is more effective in determining phylogeny.

Background on comparative genomic analysis. (2002, December). Retrieved from http://www.genome.gov/10005835

Note:

Evolution in Action

Why Does Phylogeny Matter?

In addition to enhancing our understanding of the evolutionary history of species, our own included, phylogenetic analysis has numerous practical applications. Two of those applications include understanding the evolution and transmission of disease and making decisions about conservation efforts. A 2010 study[footnote] of MRSA (methicillin-resistant *Staphylococcus aureus*), an antibiotic resistant pathogenic bacterium, traced the origin and spread of the strain throughout the past 40 years. The study uncovered the timing and patterns in which the resistant strain moved from its point of origin in Europe to centers of infection and evolution in South America, Asia, North America, and Australasia. The study suggested that introductions of the bacteria to new populations occurred very few times, perhaps only once, and then spread from that limited number of individuals. This is in contrast to the possibility that many individuals had carried the bacteria from one place to another. This result suggests that public health officials should concentrate on quickly identifying the contacts of individuals infected with a new strain of bacteria to control its spread.

Harris, S.R. et al. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474.

A second area of usefulness for phylogenetic analysis is in conservation. Biologists have argued that it is important to protect species throughout a phylogenetic tree rather than just those from one branch of the tree. Doing this will preserve more of the variation produced by evolution. For example, conservation efforts should focus on a single species without sister species rather than another species that has a cluster of close sister species that recently evolved. If the single evolutionarily distinct species

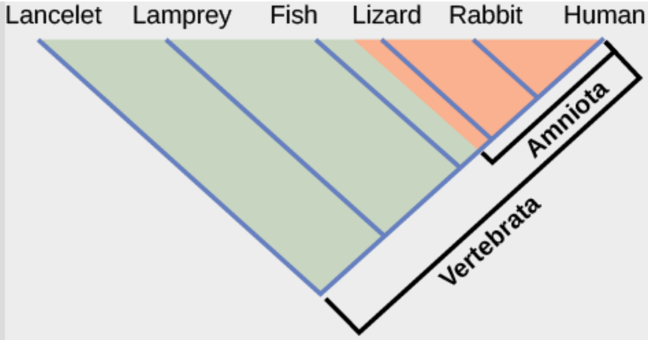
goes extinct a disproportionate amount of variation from the tree will be lost compared to one species in the cluster of closely related species. A study published in 2007[footnote] made recommendations for conservation of mammal species worldwide based on how evolutionarily distinct and at risk of extinction they are. The study found that their recommendations differed from priorities based on simply the level of extinction threat to the species. The study recommended protecting some threatened and valued large mammals such as the orangutans, the giant and lesser pandas, and the African and Asian elephants. But they also found that some much lesser known species should be protected based on how evolutionary distinct they are. These include a number of rodents, bats, shrews and hedgehogs. In addition there are some critically endangered species that did not rate as very important in evolutionary distinctiveness including species of deer mice and gerbils. While many criteria affect conservation decisions, preserving phylogenetic diversity provides an objective way to protect the full range of diversity generated by evolution.

Isaac NJ, Turvey ST, Collen B, Waterman C, Baillie JE (2007) Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. PLoS ONE 2(3): e296. doi:10.1371/journal.pone.0000296

Building Phylogenetic Trees

How do scientists construct phylogenetic trees? Presently, the most accepted method for constructing phylogenetic trees is a method called **cladistics**. This method sorts organisms into **clades**, groups of organisms that are most closely related to each other and the ancestor from which they descended. For example, in [link], all of the organisms in the shaded region evolved from a single ancestor that had amniotic eggs. Consequently, all of these organisms also have amniotic eggs and make a single clade, also called a **monophyletic group**. Clades must include the ancestral species and all of the descendants from a branch point.

Note:
Art Connection



Lizards, rabbits, and humans all descend from a common ancestor in which the amniotic egg evolved. Thus, lizards, rabbits, and humans all belong to the clade Amniota. Vertebrata is a larger clade that also includes fish and lamprey.

Which animals in this figure belong to a clade that includes animals with hair? Which evolved first: hair or the amniotic egg?

Clades can vary in size depending on which branch point is being referenced. The important factor is that all of the organisms in the clade or monophyletic group stem from a single point on the tree. This can be remembered because monophyletic breaks down into “mono,” meaning one, and “phyletic,” meaning evolutionary relationship.

Shared Characteristics

Cladistics rests on three assumptions. The first is that living things are related by descent from a common ancestor, which is a general assumption of evolution. The second is that speciation occurs by splits of one species into two, never more than two at a time, and essentially at one point in time.

This is somewhat controversial, but is acceptable to most biologists as a simplification. The third assumption is that traits change enough over time to be considered to be in a different state .It is also assumed that one can identify the actual direction of change for a state. In other words, we assume that an amniotic egg is a later character state than non-amniotic eggs. This is called the polarity of the character change. We know this by reference to a group outside the clade: for example, insects have non-amniotic eggs; therefore, this is the older or ancestral character state. Cladistics compares ingroups and outgroups. An ingroup (lizard, rabbit and human in our example) is the group of taxa being analyzed. An outgroup (lancelet, lamprey and fish in our example) is a species or group of species that diverged before the lineage containing the group(s) of interest. By comparing ingroup members to each other and to the outgroup members, we can determine which characteristics are evolutionary modifications determining the branch points of the ingroup’s phylogeny.

If a characteristic is found in all of the members of a group, it is a **shared ancestral character** because there has been no change in the trait during the descent of each of the members of the clade. Although these traits appear interesting because they unify the clade, in cladistics they are considered not helpful when we are trying to determine the relationships of the members of the clade because every member is the same. In contrast, consider the amniotic egg characteristic of [link]. Only some of the organisms have this trait, and to those that do, it is called a **shared derived character** because this trait changed at some point during descent. This character does tell us about the relationships among the members of the clade; it tells us that lizards, rabbits, and humans group more closely together than any of these organisms do with fish, lampreys, and lancelets.

A sometimes confusing aspect of “ancestral” and “derived” characters is that these terms are relative. The same trait could be either ancestral or derived depending on the diagram being used and the organisms being compared. Scientists find these terms useful when distinguishing between clades during the building of phylogenetic trees, but it is important to remember that their meaning depends on context.

Choosing the Right Relationships

Constructing a phylogenetic tree, or cladogram, from the character data is a monumental task that is usually left up to a computer. The computer draws a tree such that all of the clades share the same list of derived characters. But there are other decisions to be made, for example, what if a species presence in a clade is supported by all of the shared derived characters for that clade except one? One conclusion is that the trait evolved in the ancestor, but then changed back in that one species. Also a character state that appears in two clades must be assumed to have evolved independently in those clades. These inconsistencies are common in trees drawn from character data and complicate the decision-making process about which tree most closely represents the real relationships among the taxa.

To aid in the tremendous task of choosing the best tree, scientists often use a concept called **maximum parsimony**, which means that events occurred in the simplest, most obvious way. This means that the “best” tree is the one with the fewest number of character reversals, the fewest number of independent character changes, and the fewest number of character changes throughout the tree. Computer programs search through all of the possible trees to find the small number of trees with the simplest evolutionary pathways. Starting with all of the homologous traits in a group of organisms, scientists can determine the order of evolutionary events of which those traits occurred that is the most obvious and simple.

Note:
Concept in Action

openstax COLLEGE

Practice Parsimony: Go to this website to learn how maximum parsimony is used to create phylogenetic trees (be sure to continue to the second

page).

These tools and concepts are only a few of the strategies scientists use to tackle the task of revealing the evolutionary history of life on Earth. Recently, newer technologies have uncovered surprising discoveries with unexpected relationships, such as the fact that people seem to be more closely related to fungi than fungi are to plants. Sound unbelievable? As the information about DNA sequences grows, scientists will become closer to mapping the evolutionary history of all life on Earth.

Section Summary

To build phylogenetic trees, scientists must collect character information that allows them to make evolutionary connections between organisms. Using morphologic and molecular data, scientists work to identify homologous characteristics and genes. Similarities between organisms can stem either from shared evolutionary history (homologies) or from separate evolutionary paths (analogies). After homologous information is identified, scientists use cladistics to organize these events as a means to determine an evolutionary timeline. Scientists apply the concept of maximum parsimony, which states that the likeliest order of events is probably the simplest shortest path. For evolutionary events, this would be the path with the least number of major divergences that correlate with the evidence.

Exercise:

Problem:

[link] Which animals in this figure belong to a clade that includes animals with hair? Which evolved first: hair or the amniotic egg?

Solution:

[link] Rabbits and humans belong in the clade that includes animals with hair. The amniotic egg evolved before hair, because the Amniota clade branches off earlier than the clade that encompasses animals with hair.

Multiple Choice

Exercise:

Problem: Which statement about analogies is correct?

a. They occur only as errors.
b. They are synonymous with homologous traits.
c. They are derived by response to similar environmental pressures.
d. They are a form of mutation.

Solution:

C

Exercise:

Problem: What kind of trait is important to cladistics?

a. shared derived traits
b. shared ancestral traits
c. analogous traits
d. parsimonious traits

Solution:

A

Exercise:

Problem:

What is true about organisms that are a part of the same clade?

a. They all share the same basic characteristics.
b. They evolved from a shared ancestor.

c. They all are on the same tree.
d. They have identical phylogenies.

Solution:

B

Exercise:

Problem: Which assumption of cladistics is stated incorrectly?

a. Living things are related by descent from a common ancestor.
b. Speciation can produce one, two, or three new species.
c. Traits change from one state to another.
d. The polarity of a character state change can be determined.

Solution:

B

Exercise:

Problem: A monophyletic group is a ________.

a. phylogenetic tree
b. shared derived trait
c. character state
d. clade

Solution:

D

Free Response

Exercise:

Problem:

Dolphins and fish have similar body shapes. Is this feature more likely a homologous or analogous trait?

Solution:

Dolphins are mammals and fish are not, which means that their evolutionary paths (phylogenies) are quite separate. Dolphins probably adapted to have a similar body plan after returning to an aquatic lifestyle, and therefore this trait is probably analogous.

Exercise:

Problem: Describe maximum parsimony.

Solution:

Maximum parsimony hypothesizes that events occurred in the simplest, most obvious way, and the pathway of evolution probably includes the fewest major events that coincide with the evidence at hand.

Exercise:

Problem:

How does a biologist determine the polarity of a character change?

Solution:

The biologist looks at the state of the character in an outgroup, an organism that is outside the clade for which the phylogeny is being developed. The polarity of the character change is from the state of the character in the outgroup to the second state.

Glossary

analogous structure
: a character found in two taxa that looks similar because of convergent evolution, not because of descent from a common ancestor

clade
: a group of taxa with the same set of shared derived characters, including an ancestral species and all its descendants

cladistics
: a method used to organize homologous traits to describe phylogenies using common descendent as the primary criterion used to classify organisms

maximum parsimony
: applying the simplest, most obvious way with the least number of steps

molecular systematics
: the methods of using molecular evidence to identify phylogenetic relationships

monophyletic group
: (also, clade) organisms that share a single ancestor

shared ancestral character
: a character on a phylogenetic branch that is shared by a particular clade

shared derived character
: a character on a phylogenetic tree that is shared only by a certain clade of organisms

Viral Evolution, Morphology, and Classification
By the end of this section, you will be able to:

- Describe how viruses were first discovered and how they are detected
- Discuss three hypotheses about how viruses evolved
- Recognize the basic shapes of viruses
- Understand basic classification systems for viruses

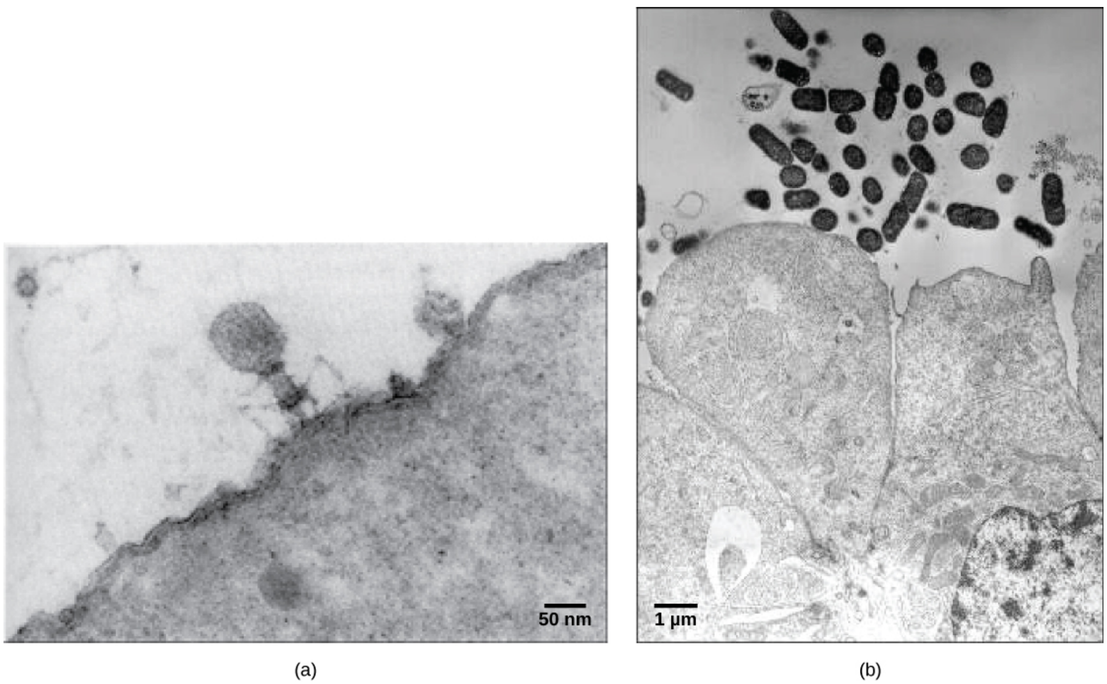
Viruses are diverse entities. They vary in their structure, their replication methods, and in their target hosts. Nearly all forms of life—from bacteria and archaea to eukaryotes such as plants, animals, and fungi—have viruses that infect them. While most biological diversity can be understood through evolutionary history, such as how species have adapted to conditions and environments, much about virus origins and evolution remains unknown.

Discovery and Detection

Viruses were first discovered after the development of a porcelain filter, called the Chamberland-Pasteur filter, which could remove all bacteria visible in the microscope from any liquid sample. In 1886, Adolph Meyer demonstrated that a disease of tobacco plants, tobacco mosaic disease, could be transferred from a diseased plant to a healthy one via liquid plant extracts. In 1892, Dmitri Ivanowski showed that this disease could be transmitted in this way even after the Chamberland-Pasteur filter had removed all viable bacteria from the extract. Still, it was many years before it was proven that these “filterable” infectious agents were not simply very small bacteria but were a new type of very small, disease-causing particle.

Virions, single virus particles, are very small, about 20–250 nanometers in diameter. These individual virus particles are the infectious form of a virus outside the host cell. Unlike bacteria (which are about 100-times larger), we cannot see viruses with a light microscope, with the exception of some large virions of the poxvirus family. It was not until the development of the electron microscope in the late 1930s that scientists got their first good view of the structure of the tobacco mosaic virus (TMV) ([link]) and other viruses ([link]). The surface structure of virions can be observed by both scanning and transmission electron microscopy, whereas the internal

structures of the virus can only be observed in images from a transmission electron microscope. The use of these technologies has allowed for the discovery of many viruses of all types of living organisms. They were initially grouped by shared morphology. Later, groups of viruses were classified by the type of nucleic acid they contained, DNA or RNA, and whether their nucleic acid was single- or double-stranded. More recently, molecular analysis of viral replicative cycles has further refined their classification.



In these transmission electron micrographs, (a) a virus is dwarfed by the bacterial cell it infects, while (b) these *E. coli* cells are dwarfed by cultured colon cells. (credit a: modification of work by U.S. Dept. of Energy, Office of Science, LBL, PBD; credit b: modification of work by J.P. Nataro and S. Sears, unpub. data, CDC; scale-bar data from Matt Russell)

Evolution of Viruses

Although biologists have accumulated a significant amount of knowledge about how present-day viruses evolve, much less is known about how viruses originated in the first place. When exploring the evolutionary history of most organisms, scientists can look at fossil records and similar historic evidence. However, viruses do not fossilize, so researchers must conjecture by investigating how today's viruses evolve and by using biochemical and genetic information to create speculative virus histories.

While most findings agree that viruses don't have a single common ancestor, scholars have yet to find a single hypothesis about virus origins that is fully accepted in the field. One such hypothesis, called devolution or the regressive hypothesis, proposes to explain the origin of viruses by suggesting that viruses evolved from free-living cells. However, many components of how this process might have occurred are a mystery. A second hypothesis (called escapist or the progressive hypothesis) accounts for viruses having either an RNA or a DNA genome and suggests that viruses originated from RNA and DNA molecules that escaped from a host cell. A third hypothesis posits a system of self-replication similar to that of other self-replicating molecules, likely evolving alongside the cells they rely on as hosts; studies of some plant pathogens support this hypothesis.

As technology advances, scientists may develop and refine further hypotheses to explain the origin of viruses. The emerging field called virus molecular systematics attempts to do just that through comparisons of sequenced genetic material. These researchers hope to one day better understand the origin of viruses, a discovery that could lead to advances in the treatments for the ailments they produce.

Viral Morphology

Viruses are **acellular**, meaning they are biological entities that do not have a cellular structure. They therefore lack most of the components of cells, such as organelles, ribosomes, and the plasma membrane. A virion consists of a nucleic acid core, an outer protein coating or capsid, and sometimes an outer **envelope** made of protein and phospholipid membranes derived from

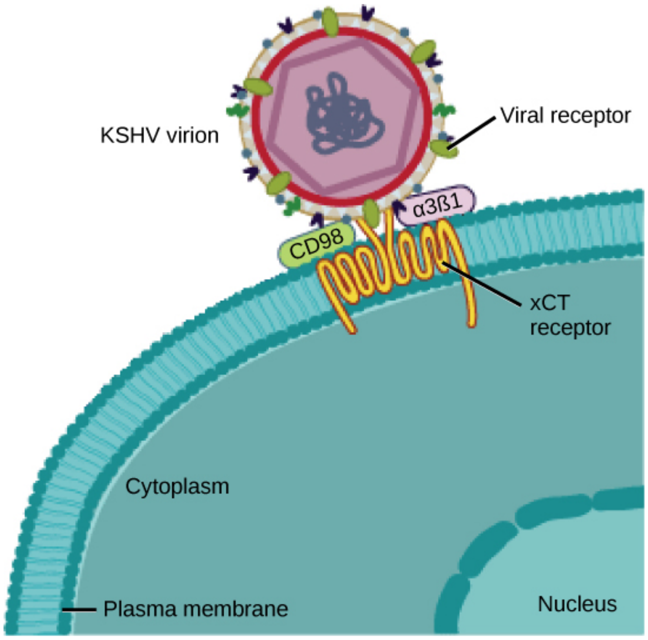
the host cell. Viruses may also contain additional proteins, such as enzymes. The most obvious difference between members of viral families is their morphology, which is quite diverse. An interesting feature of viral complexity is that the complexity of the host does not correlate with the complexity of the virion. Some of the most complex virion structures are observed in bacteriophages, viruses that infect the simplest living organisms, bacteria.

Morphology

Viruses come in many shapes and sizes, but these are consistent and distinct for each viral family. All virions have a nucleic acid genome covered by a protective layer of proteins, called a **capsid**. The capsid is made up of protein subunits called **capsomeres**. Some viral capsids are simple polyhedral “spheres,” whereas others are quite complex in structure.

In general, the shapes of viruses are classified into four groups: filamentous, isometric (or icosahedral), enveloped, and head and tail. Filamentous viruses are long and cylindrical. Many plant viruses are filamentous, including TMV. Isometric viruses have shapes that are roughly spherical, such as poliovirus or herpesviruses. Enveloped viruses have membranes surrounding capsids. Animal viruses, such as HIV, are frequently enveloped. Head and tail viruses infect bacteria and have a head that is similar to icosahedral viruses and a tail shape like filamentous viruses.

Many viruses use some sort of glycoprotein to attach to their host cells via molecules on the cell called **viral receptors** ([link]). For these viruses, attachment is a requirement for later penetration of the cell membrane, so they can complete their replication inside the cell. The receptors that viruses use are molecules that are normally found on cell surfaces and have their own physiological functions. Viruses have simply evolved to make use of these molecules for their own replication. For example, HIV uses the CD4 molecule on T lymphocytes as one of its receptors. CD4 is a type of molecule called a cell adhesion molecule, which functions to keep different types of immune cells in close proximity to each other during the generation of a T lymphocyte immune response.



The KSHV virus binds the xCT receptor on the surface of human cells. xCT receptors protect cells against stress. Stressed cells express more xCT receptors than non-stressed cells. The KSHV virion causes cells to become stressed, thereby increasing expression of the receptor to which it binds. (credit: modification of work by NIAID, NIH)

Among the most complex virions known, the T4 bacteriophage, which infects the *Escherichia coli* bacterium, has a tail structure that the virus uses to attach to host cells and a head structure that houses its DNA.

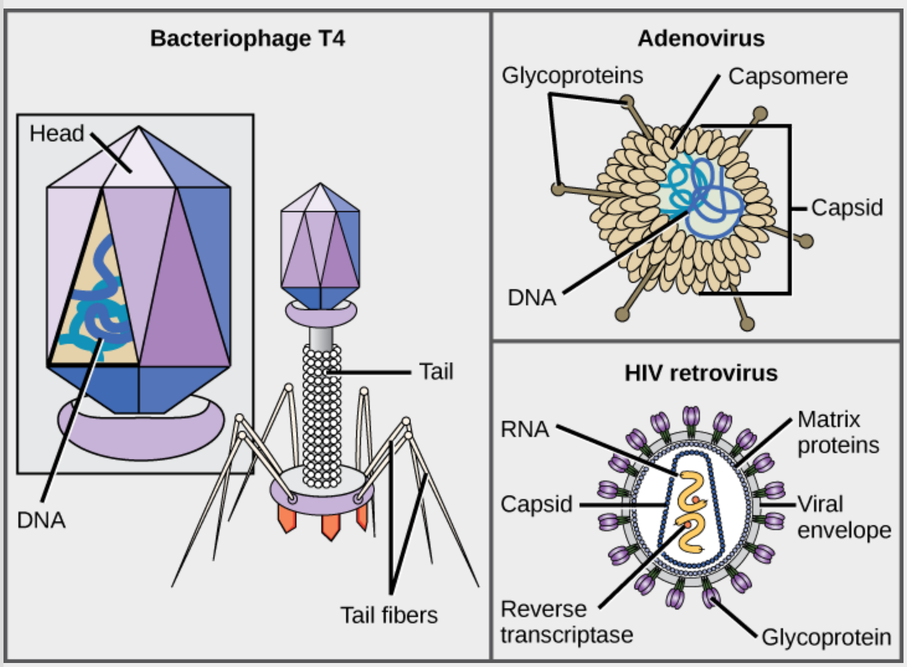
Adenovirus, a non-enveloped animal virus that causes respiratory illnesses in humans, uses glycoprotein spikes protruding from its capsomeres to attach to host cells. Non-enveloped viruses also include those that cause

polio (poliovirus), plantar warts (papillomavirus), and hepatitis A (hepatitis A virus).

Enveloped virions like HIV, the causative agent in AIDS, consist of nucleic acid (RNA in the case of HIV) and capsid proteins surrounded by a phospholipid bilayer envelope and its associated proteins. Glycoproteins embedded in the viral envelope are used to attach to host cells. Other envelope proteins are the **matrix proteins** that stabilize the envelope and often play a role in the assembly of progeny virions. Chicken pox, influenza, and mumps are examples of diseases caused by viruses with envelopes. Because of the fragility of the envelope, non-enveloped viruses are more resistant to changes in temperature, pH, and some disinfectants than enveloped viruses.

Overall, the shape of the virion and the presence or absence of an envelope tell us little about what disease the virus may cause or what species it might infect, but they are still useful means to begin viral classification ([link]).

Note:
Art Connection



Viruses can be either complex in shape or relatively simple. This figure shows three relatively complex virions: the bacteriophage T4, with its DNA-containing head group and tail fibers that attach to host cells; adenovirus, which uses spikes from its capsid to bind to host cells; and HIV, which uses glycoproteins embedded in its envelope to bind to host cells. Notice that HIV has proteins called matrix proteins, internal to the envelope, which help stabilize virion shape. (credit “bacteriophage, adenovirus”: modification of work by NCBI, NIH; credit “HIV retrovirus”: modification of work by NIAID, NIH)

Which of the following statements about virus structure is true?

a. All viruses are encased in a viral membrane.
b. The capsomere is made up of small protein subunits called capsids.
c. DNA is the genetic material in all viruses.
d. Glycoproteins help the virus attach to the host cell.

Types of Nucleic Acid

Unlike nearly all living organisms that use DNA as their genetic material, viruses may use either DNA or RNA as theirs. The **virus core** contains the genome or total genetic content of the virus. Viral genomes tend to be small, containing only those genes that encode proteins that the virus cannot get from the host cell. This genetic material may be single- or double-stranded. It may also be linear or circular. While most viruses contain a single nucleic acid, others have genomes that have several, which are called segments.

In DNA viruses, the viral DNA directs the host cell’s replication proteins to synthesize new copies of the viral genome and to transcribe and translate

that genome into viral proteins. DNA viruses cause human diseases, such as chickenpox, hepatitis B, and some venereal diseases, like herpes and genital warts.

RNA viruses contain only RNA as their genetic material. To replicate their genomes in the host cell, the RNA viruses encode enzymes that can replicate RNA into DNA, which cannot be done by the host cell. These RNA polymerase enzymes are more likely to make copying errors than DNA polymerases, and therefore often make mistakes during transcription. For this reason, mutations in RNA viruses occur more frequently than in DNA viruses. This causes them to change and adapt more rapidly to their host. Human diseases caused by RNA viruses include hepatitis C, measles, and rabies.

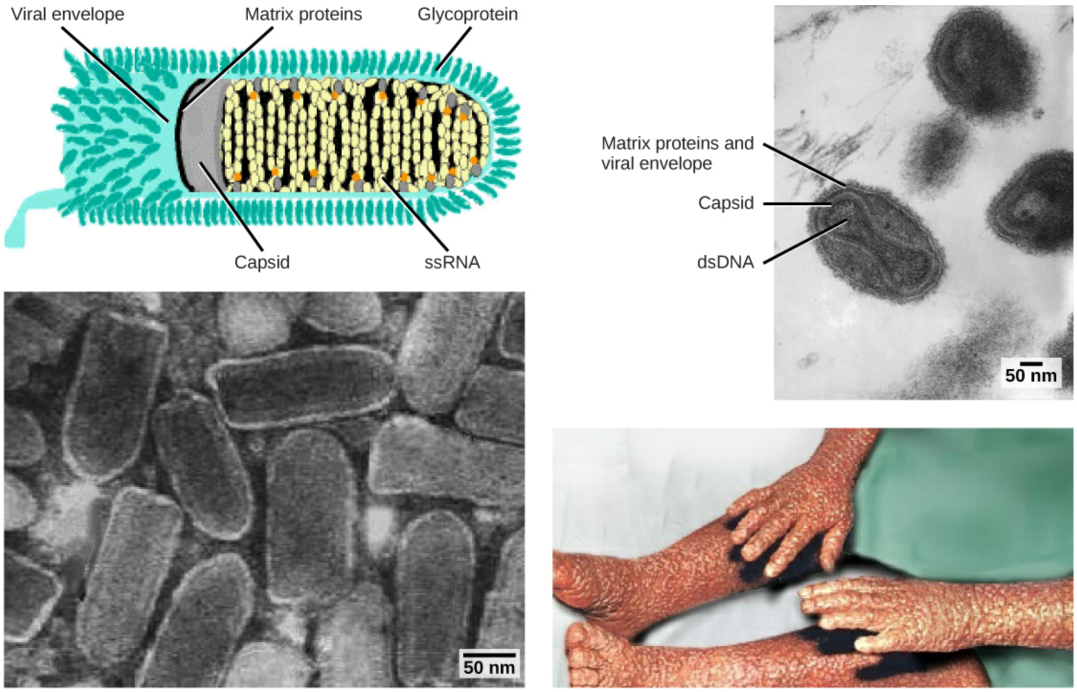
Virus Classification

To understand the features shared among different groups of viruses, a classification scheme is necessary. As most viruses are not thought to have evolved from a common ancestor, however, the methods that scientists use to classify living things are not very useful. Biologists have used several classification systems in the past, based on the morphology and genetics of the different viruses. However, these earlier classification methods grouped viruses differently, based on which features of the virus they were using to classify them. The most commonly used classification method today is called the Baltimore classification scheme and is based on how messenger RNA (mRNA) is generated in each particular type of virus.

Past Systems of Classification

Viruses are classified in several ways: by factors such as their core content ([link] and [link]), the structure of their capsids, and whether they have an outer envelope. The type of genetic material (DNA or RNA) and its structure (single- or double-stranded, linear or circular, and segmented or non-segmented) are used to classify the virus core structures.

Virus Classification by Genome Structure and Core	
Core Classifications	**Examples**
• RNA • DNA	• Rabies virus, retroviruses • Herpesviruses, smallpox virus
• Single-stranded • Double-stranded	• Rabies virus, retroviruses • Herpesviruses, smallpox virus
• Linear • Circular	Rabies virus, retroviruses, herpesviruses, smallpox virus Papillomaviruses, many bacteriophages
• Non-segmented: genome consists of a single segment of genetic material • Segmented: genome is divided into multiple segments	Parainfluenza viruses Influenza viruses

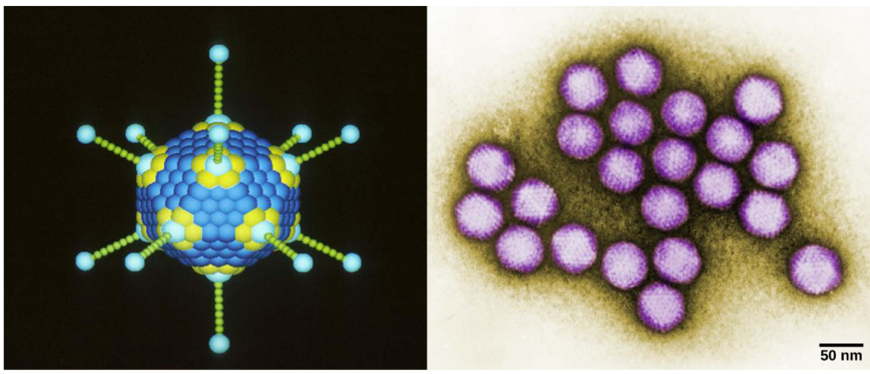


(a) Rabies virus

(b) Variola virus

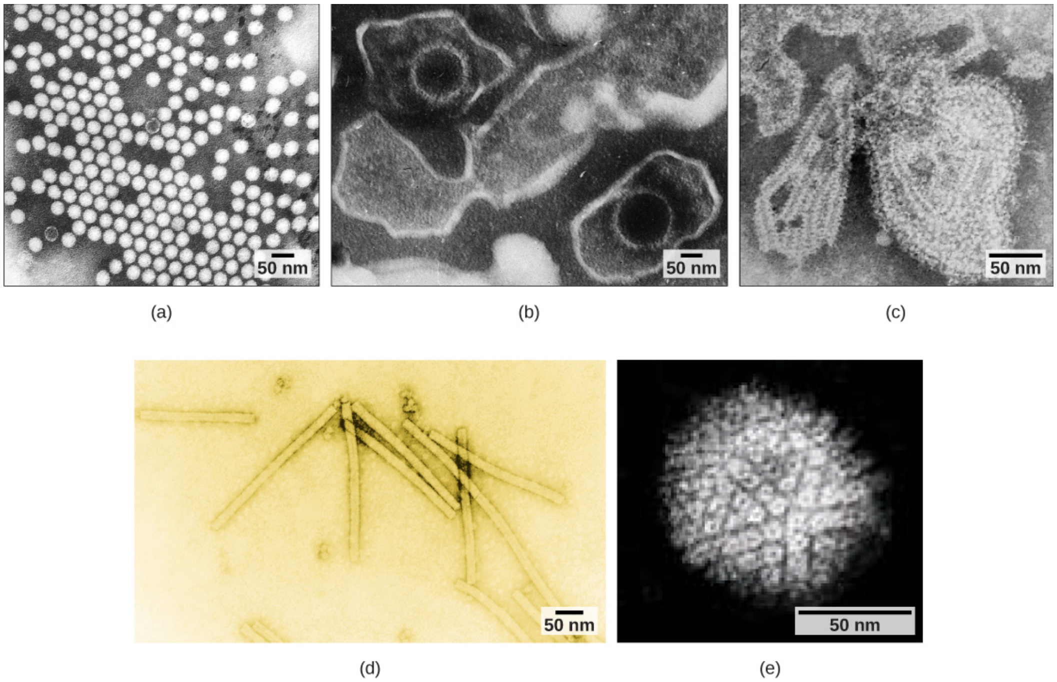
Viruses are classified based on their core genetic material and capsid design. (a) Rabies virus has a single-stranded RNA (ssRNA) core and an enveloped helical capsid, whereas (b) variola virus, the causative agent of smallpox, has a double-stranded DNA (dsDNA) core and a complex capsid. Rabies transmission occurs when saliva from an infected mammal enters a wound. The virus travels through neurons in the peripheral nervous system to the central nervous system where it impairs brain function, and then travels to other tissues. The virus can infect any mammal, and most die within weeks of infection. Smallpox is a human virus transmitted by inhalation of the variola virus, localized in the skin, mouth, and throat, which causes a characteristic rash. Before its eradication in 1979, infection resulted in a 30–35 percent mortality rate. (credit “rabies diagram”: modification of work by CDC; “rabies micrograph”: modification of work by Dr. Fred Murphy, CDC; credit “small pox micrograph”: modification of work by Dr. Fred Murphy, Sylvia Whitfield, CDC; credit “smallpox photo”: modification of work by CDC; scale-bar data from Matt Russell)

Viruses can also be classified by the design of their capsids ([link] and [link]). Capsids are classified as naked icosahedral, enveloped icosahedral, enveloped helical, naked helical, and complex ([link] and [link]). The type of genetic material (DNA or RNA) and its structure (single- or double-stranded, linear or circular, and segmented or non-segmented) are used to classify the virus core structures ([link]).



Adenovirus (left) is depicted with a double-stranded DNA genome enclosed in an icosahedral capsid that is 90–100 nm across. The virus, shown clustered in the micrograph (right), is transmitted orally and causes a variety of illnesses in vertebrates, including human eye and respiratory infections. (credit “adenovirus”: modification of work by Dr. Richard Feldmann, National Cancer Institute; credit “micrograph”: modification of work by Dr. G. William Gary, Jr., CDC; scale-bar data from Matt Russell)

Virus Classification by Capsid Structure	
Capsid Classification	**Examples**
Naked icosahedral	Hepatitis A virus, polioviruses
Enveloped icosahedral	Epstein-Barr virus, herpes simplex virus, rubella virus, yellow fever virus, HIV-1
Enveloped helical	Influenza viruses, mumps virus, measles virus, rabies virus
Naked helical	Tobacco mosaic virus
Complex with many proteins; some have combinations of icosahedral and helical capsid structures	Herpesviruses, smallpox virus, hepatitis B virus, T4 bacteriophage



Transmission electron micrographs of various viruses show their structures. The capsid of the (a) polio virus is naked icosahedral; (b) the Epstein-Barr virus capsid is enveloped icosahedral; (c) the mumps virus capsid is an enveloped helix; (d) the tobacco mosaic virus capsid is naked helical; and (e) the herpesvirus capsid is complex. (credit a: modification of work by Dr. Fred Murphy, Sylvia Whitfield; credit b: modification of work by Liza Gross; credit c: modification of work by Dr. F. A. Murphy, CDC; credit d: modification of work by USDA ARS; credit e: modification of work by Linda Stannard, Department of Medical Microbiology, University of Cape Town, South Africa, NASA; scale-bar data from Matt Russell)

Section Summary

Viruses are tiny, acellular entities that can usually only be seen with an electron microscope. Their genomes contain either DNA or RNA—never both—and they replicate using the replication proteins of a host cell.

Viruses are diverse, infecting archaea, bacteria, fungi, plants, and animals. Viruses consist of a nucleic acid core surrounded by a protein capsid with or without an outer lipid envelope. The capsid shape, presence of an envelope, and core composition dictate some elements of the classification of viruses. The most commonly used classification method, the Baltimore classification, categorizes viruses based on how they produce their mRNA.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about virus structure is true?

a. All viruses are encased in a viral membrane.
b. The capsomere is made up of small protein subunits called capsids.
c. DNA is the genetic material in all viruses.
d. Glycoproteins help the virus attach to the host cell.

Solution:

[link] D

Review Questions

Exercise:

Problem: Which statement is true?

a. A virion contains DNA and RNA.
b. Viruses are acellular.
c. Viruses replicate outside of the cell.
d. Most viruses are easily visualized with a light microscope.

Solution:

B

Exercise:

Problem:

The viral ________ plays a role in attaching a virion to the host cell.

a. core
b. capsid
c. envelope
d. both b and c

Solution:

D

Exercise:

Problem: Viruses_______.

a. all have a round shape
b. cannot have a long shape
c. do not maintain any shape
d. vary in shape

Solution:

D

Free Response

Exercise:

Problem:

The first electron micrograph of a virus (tobacco mosaic virus) was produced in 1939. Before that time, how did scientists know that viruses existed if they could not see them? (Hint: Early scientists called viruses “filterable agents.”)

Solution:

Viruses pass through filters that eliminated all bacteria that were visible in the light microscopes at the time. As the bacteria-free filtrate could still cause infections when given to a healthy organism, this observation demonstrated the existence of very small infectious agents. These agents were later shown to be unrelated to bacteria and were classified as viruses.

Glossary

acellular
: lacking cells

capsid
: protein coating of the viral core

capsomere
: protein subunit that makes up the capsid

envelope
: lipid bilayer that envelopes some viruses

group I virus
: virus with a dsDNA genome

group II virus
: virus with a ssDNA genome

group III virus

virus with a dsRNA genome

group IV virus
virus with a ssRNA genome with positive polarity

group V virus
virus with a ssRNA genome with negative polarity

group VI virus
virus with a ssRNA genomes converted into dsDNA by reverse transcriptase

group VII virus
virus with a single-stranded mRNA converted into dsDNA for genome replication

matrix protein
envelope protein that stabilizes the envelope and often plays a role in the assembly of progeny virions

negative polarity
ssRNA viruses with genomes complimentary to their mRNA

positive polarity
ssRNA virus with a genome that contains the same base sequences and codons found in their mRNA

replicative intermediate
dsRNA intermediate made in the process of copying genomic RNA

reverse transcriptase
enzyme found in Baltimore groups VI and VII that converts single-stranded RNA into double-stranded DNA

viral receptor
glycoprotein used to attach a virus to host cells via molecules on the cell

virion

individual virus particle outside a host cell

virus core

contains the virus genome

Virus Infections and Hosts
By the end of this section, you will be able to:

- List the steps of replication and explain what occurs at each step
- Describe the lytic and lysogenic cycles of virus replication
- Explain the transmission and diseases of animal and plant viruses
- Discuss the economic impact of animal and plant viruses

Viruses can be seen as obligate, intracellular parasites. A virus must attach to a living cell, be taken inside, manufacture its proteins and copy its genome, and find a way to escape the cell so that the virus can infect other cells. Viruses can infect only certain species of hosts and only certain cells within that host. Cells that a virus may use to replicate are called **permissive**. For most viruses, the molecular basis for this specificity is that a particular surface molecule known as the viral receptor must be found on the host cell surface for the virus to attach. Also, metabolic and host cell immune response differences seen in different cell types based on differential gene expression are a likely factor in which cells a virus may target for replication. The permissive cell must make the substances that the virus needs or the virus will not be able to replicate there.

Steps of Virus Infections

A virus must use cell processes to replicate. The viral replication cycle can produce dramatic biochemical and structural changes in the host cell, which may cause cell damage. These changes, called **cytopathic** (causing cell damage) effects, can change cell functions or even destroy the cell. Some infected cells, such as those infected by the common cold virus known as rhinovirus, die through **lysis** (bursting) or apoptosis (programmed cell death or “cell suicide”), releasing all progeny virions at once. The symptoms of viral diseases result from the immune response to the virus, which attempts to control and eliminate the virus from the body, and from cell damage caused by the virus. Many animal viruses, such as HIV (human immunodeficiency virus), leave the infected cells of the immune system by a process known as **budding**, where virions leave the cell individually. During the budding process, the cell does not undergo lysis and is not immediately killed. However, the damage to the cells that the virus infects

may make it impossible for the cells to function normally, even though the cells remain alive for a period of time. Most productive viral infections follow similar steps in the virus replication cycle: attachment, penetration, uncoating, replication, assembly, and release ([link]).

Attachment

A virus attaches to a specific receptor site on the host cell membrane through attachment proteins in the capsid or via glycoproteins embedded in the viral envelope. The specificity of this interaction determines the host—and the cells within the host—that can be infected by a particular virus. This can be illustrated by thinking of several keys and several locks, where each key will fit only one specific lock.

Note:
Link to Learning

This video explains how influenza attacks the body.

Entry

The nucleic acid of bacteriophages enters the host cell naked, leaving the capsid outside the cell. Plant and animal viruses can enter through endocytosis, in which the cell membrane surrounds and engulfs the entire virus. Some enveloped viruses enter the cell when the viral envelope fuses directly with the cell membrane. Once inside the cell, the viral capsid is

degraded, and the viral nucleic acid is released, which then becomes available for replication and transcription.

Replication and Assembly

The replication mechanism depends on the viral genome. DNA viruses usually use host cell proteins and enzymes to make additional DNA that is transcribed to messenger RNA (mRNA), which is then used to direct protein synthesis. RNA viruses usually use the RNA core as a template for synthesis of viral genomic RNA and mRNA. The viral mRNA directs the host cell to synthesize viral enzymes and capsid proteins, and assemble new virions. Of course, there are exceptions to this pattern. If a host cell does not provide the enzymes necessary for viral replication, viral genes supply the information to direct synthesis of the missing proteins. Retroviruses, such as HIV, have an RNA genome that must be reverse transcribed into DNA, which then is incorporated into the host cell genome. They are within group VI of the Baltimore classification scheme. To convert RNA into DNA, retroviruses must contain genes that encode the virus-specific enzyme reverse transcriptase that transcribes an RNA template to DNA. Reverse transcription never occurs in uninfected host cells—the needed enzyme reverse transcriptase is only derived from the expression of viral genes within the infected host cells. The fact that HIV produces some of its own enzymes not found in the host has allowed researchers to develop drugs that inhibit these enzymes. These drugs, including the reverse transcriptase inhibitor **AZT**, inhibit HIV replication by reducing the activity of the enzyme without affecting the host's metabolism. This approach has led to the development of a variety of drugs used to treat HIV and has been effective at reducing the number of infectious virions (copies of viral RNA) in the blood to non-detectable levels in many HIV-infected individuals.

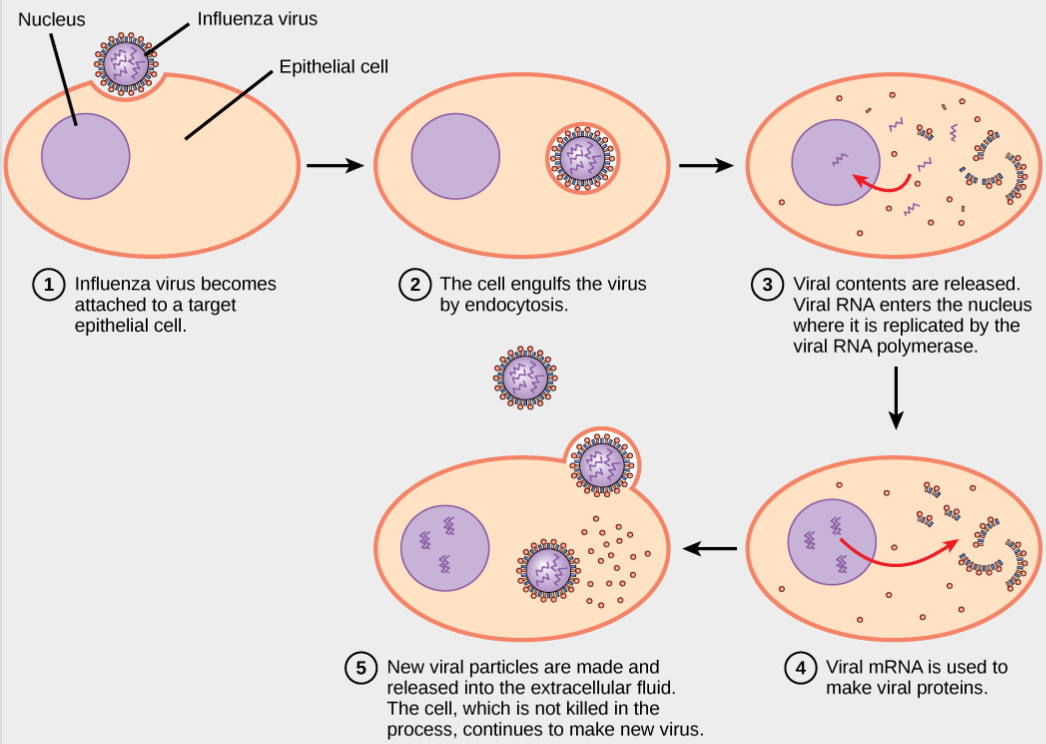
Egress

The last stage of viral replication is the release of the new virions produced in the host organism, where they are able to infect adjacent cells and repeat

the replication cycle. As you’ve learned, some viruses are released when the host cell dies, and other viruses can leave infected cells by budding through the membrane without directly killing the cell.

Note:

Art Connection



In influenza virus infection, glycoproteins attach to a host epithelial cell. As a result, the virus is engulfed. RNA and proteins are made and assembled into new virions.

Influenza virus is packaged in a viral envelope that fuses with the plasma membrane. This way, the virus can exit the host cell without killing it. What advantage does the virus gain by keeping the host cell alive?

Note:

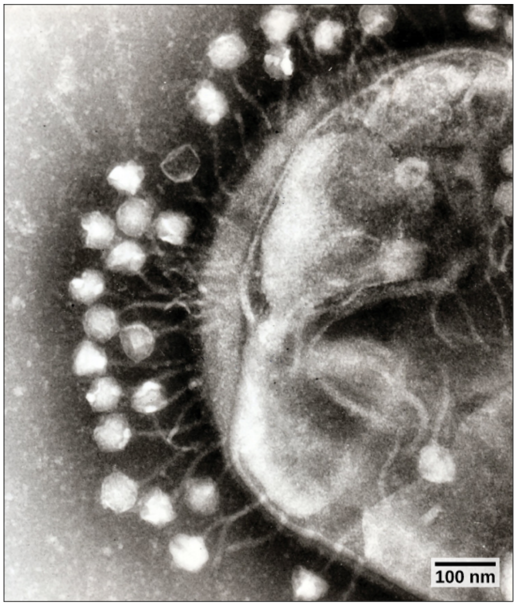
Link to Learning

Watch a video on viruses, identifying structures, modes of transmission, replication, and more.

Different Hosts and Their Viruses

As you’ve learned, viruses are often very specific as to which hosts and which cells within the host they will infect. This feature of a virus makes it specific to one or a few species of life on Earth. On the other hand, so many different types of viruses exist on Earth that nearly every living organism has its own set of viruses that tries to infect its cells. Even the smallest and simplest of cells, prokaryotic bacteria, may be attacked by specific types of viruses.

Bacteriophages



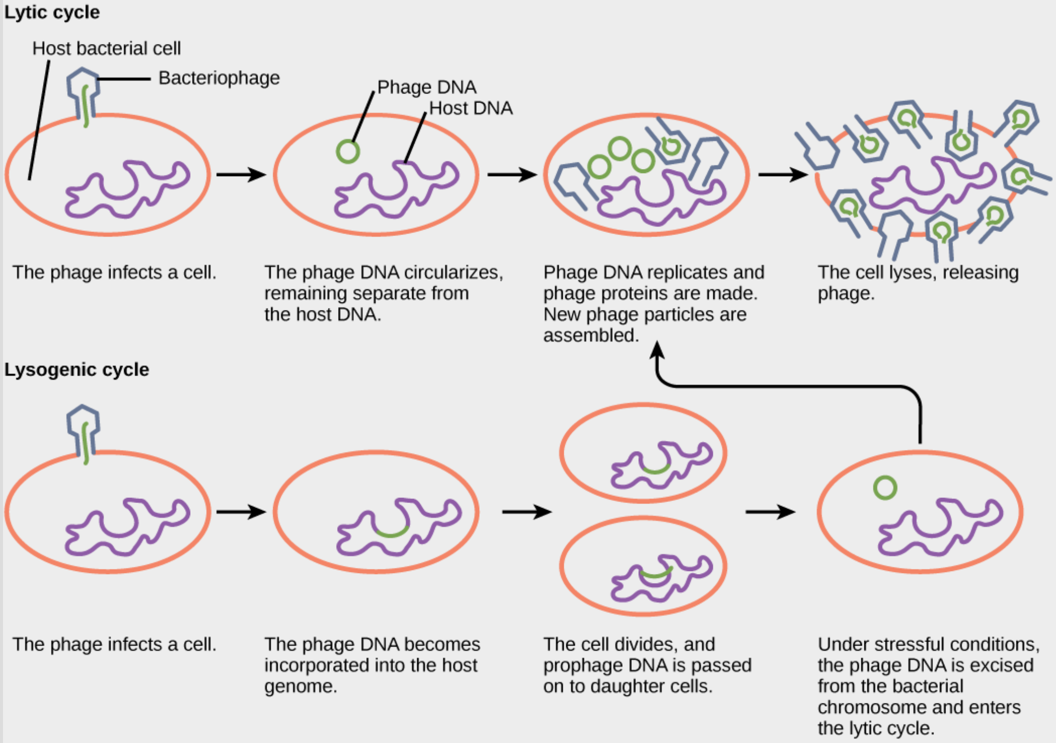
This transmission electron micrograph shows bacteriophages attached to a bacterial cell. (credit: modification of work by Dr. Graham Beards; scale-bar data from Matt Russell)

Bacteriophages are viruses that infect bacteria ([link]). When infection of a cell by a bacteriophage results in the production of new virions, the infection is said to be **productive**. If the virions are released by bursting the cell, the virus replicates by means of a **lytic cycle** ([link]). An example of a lytic bacteriophage is T4, which infects *Escherichia coli* found in the human intestinal tract. Sometimes, however, a virus can remain within the cell without being released. For example, when a temperate bacteriophage infects a bacterial cell, it replicates by means of a **lysogenic cycle** ([link]), and the viral genome is incorporated into the genome of the host cell. When the phage DNA is incorporated into the host cell genome, it is called a **prophage**. An example of a lysogenic bacteriophage is the λ (lambda) virus, which also infects the *E. coli* bacterium. Viruses that infect plant or

animal cells may also undergo infections where they are not producing virions for long periods. An example is the animal herpesviruses, including herpes simplex viruses, the cause of oral and genital herpes in humans. In a process called **latency**, these viruses can exist in nervous tissue for long periods of time without producing new virions, only to leave latency periodically and cause lesions in the skin where the virus replicates. Even though there are similarities between lysogeny and latency, the term lysogenic cycle is usually reserved to describe bacteriophages. Latency will be described in more detail below.

Note:

Art Connection



A temperate bacteriophage has both lytic and lysogenic cycles. In the lytic cycle, the phage replicates and lyses the host cell. In the lysogenic cycle, phage DNA is incorporated into the host genome, where it is passed on to subsequent generations. Environmental stressors such as starvation or

exposure to toxic chemicals may cause the prophage to excise and enter the lytic cycle.

Which of the following statements is false?

a. In the lytic cycle, new phage are produced and released into the environment.
b. In the lysogenic cycle, phage DNA is incorporated into the host genome.
c. An environmental stressor can cause the phage to initiate the lysogenic cycle.
d. Cell lysis only occurs in the lytic cycle.

Animal Viruses

Animal viruses, unlike the viruses of plants and bacteria, do not have to penetrate a cell wall to gain access to the host cell. Non-enveloped or "naked" animal viruses may enter cells in two different ways. As a protein in the viral capsid binds to its receptor on the host cell, the virus may be taken inside the cell via a vesicle during the normal cell process of receptor-mediated endocytosis. An alternative method of cell penetration used by non-enveloped viruses is for capsid proteins to undergo shape changes after binding to the receptor, creating channels in the host cell membrane. The viral genome is then "injected" into the host cell through these channels in a manner analogous to that used by many bacteriophages. Enveloped viruses also have two ways of entering cells after binding to their receptors: receptor-mediated endocytosis, or **fusion**. Many enveloped viruses enter the cell by receptor-mediated endocytosis in a fashion similar to some non-enveloped viruses. On the other hand, fusion only occurs with enveloped virions. These viruses, which include HIV among others, use special fusion proteins in their envelopes to cause the envelope to fuse with the plasma membrane of the cell, thus releasing the genome and capsid of the virus into the cell cytoplasm.

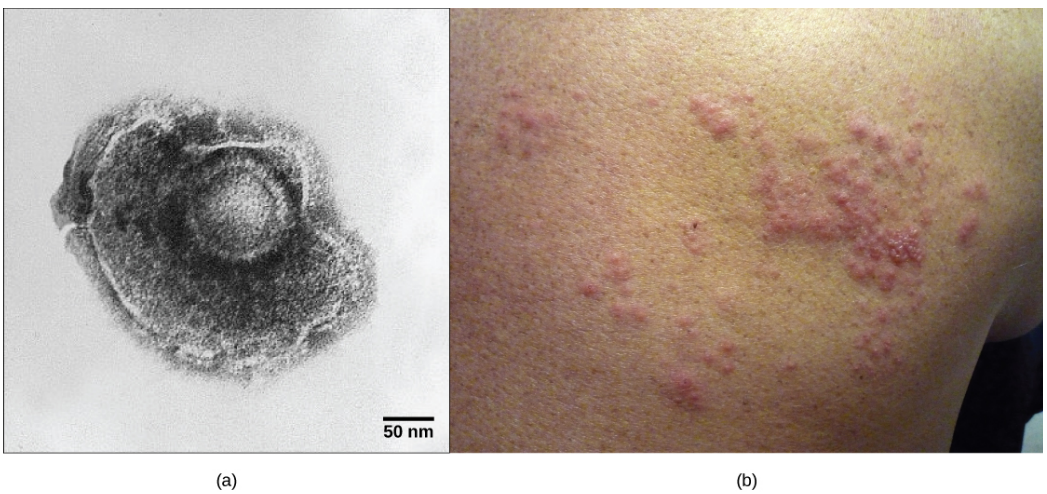
After making their proteins and copying their genomes, animal viruses complete the assembly of new virions and exit the cell. As we have already discussed using the example of HIV, enveloped animal viruses may bud from the cell membrane as they assemble themselves, taking a piece of the cell's plasma membrane in the process. On the other hand, non-enveloped viral progeny, such as rhinoviruses, accumulate in infected cells until there is a signal for lysis or apoptosis, and all virions are released together.

As you will learn in the next module, animal viruses are associated with a variety of human diseases. Some of them follow the classic pattern of **acute disease**, where symptoms get increasingly worse for a short period followed by the elimination of the virus from the body by the immune system and eventual recovery from the infection. Examples of acute viral diseases are the common cold and influenza. Other viruses cause long-term **chronic infections**, such as the virus causing hepatitis C, whereas others, like herpes simplex virus, only cause **intermittent** symptoms. Still other viruses, such as human herpesviruses 6 and 7, which in some cases can cause the minor childhood disease roseola, often successfully cause productive infections without causing any symptoms at all in the host, and thus we say these patients have an **asymptomatic infection**.

In hepatitis C infections, the virus grows and reproduces in liver cells, causing low levels of liver damage. The damage is so low that infected individuals are often unaware that they are infected, and many infections are detected only by routine blood work on patients with risk factors such as intravenous drug use. On the other hand, since many of the symptoms of viral diseases are caused by immune responses, a lack of symptoms is an indication of a weak immune response to the virus. This allows for the virus to escape elimination by the immune system and persist in individuals for years, all the while producing low levels of progeny virions in what is known as a chronic viral disease. Chronic infection of the liver by this virus leads to a much greater chance of developing liver cancer, sometimes as much as 30 years after the initial infection.

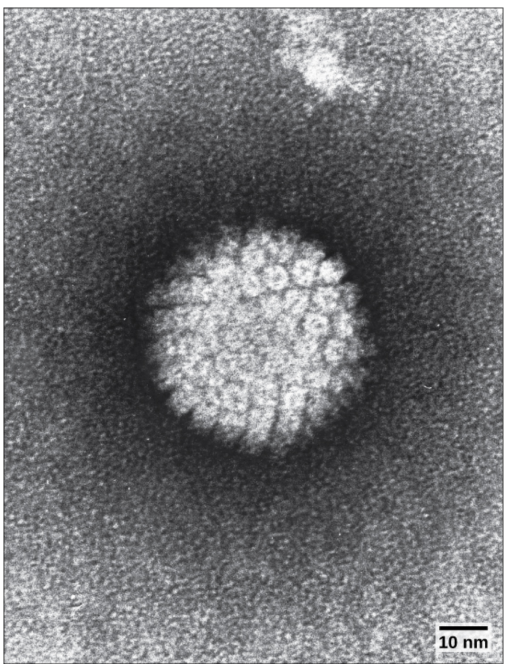
As already discussed, herpes simplex virus can remain in a state of latency in nervous tissue for months, even years. As the virus "hides" in the tissue and makes few if any viral proteins, there is nothing for the immune

response to act against, and immunity to the virus slowly declines. Under certain conditions, including various types of physical and psychological stress, the latent herpes simplex virus may be reactivated and undergo a lytic replication cycle in the skin, causing the lesions associated with the disease. Once virions are produced in the skin and viral proteins are synthesized, the immune response is again stimulated and resolves the skin lesions in a few days by destroying viruses in the skin. As a result of this type of replicative cycle, appearances of cold sores and genital herpes outbreaks only occur intermittently, even though the viruses remain in the nervous tissue for life. Latent infections are common with other herpesviruses as well, including the varicella-zoster virus that causes chickenpox. After having a chickenpox infection in childhood, the varicella-zoster virus can remain latent for many years and reactivate in adults to cause the painful condition known as “shingles” ([link]**ab**).



(a) Varicella-zoster, the virus that causes chickenpox, has an enveloped icosahedral capsid visible in this transmission electron micrograph. Its double-stranded DNA genome becomes incorporated in the host DNA and can reactivate after latency in the form of (b) shingles, often exhibiting a rash. (credit a: modification of work by Dr. Erskine Palmer, B. G. Martin, CDC; credit b: modification of work by “rosmary”/Flickr; scale-bar data from Matt Russell)

Some animal-infecting viruses, including the hepatitis C virus discussed above, are known as **oncogenic viruses**: They have the ability to cause cancer. These viruses interfere with the normal regulation of the host cell cycle either by either introducing genes that stimulate unregulated cell growth (oncogenes) or by interfering with the expression of genes that inhibit cell growth. Oncogenic viruses can be either DNA or RNA viruses. Cancers known to be associated with viral infections include cervical cancer caused by human papillomavirus (HPV) ([link]), liver cancer caused by hepatitis B virus, T-cell leukemia, and several types of lymphoma.



HPV, or human papillomavirus, has a naked icosahedral capsid visible in this transmission electron micrograph and a double-stranded DNA genome that is incorporated into

the host DNA. The virus, which is sexually transmitted, is oncogenic and can lead to cervical cancer. (credit: modification of work by NCI, NIH; scale-bar data from Matt Russell)

Note:
Link to Learning

Visit the interactive animations showing the various stages of the replicative cycles of animal viruses and click on the flash animation links.

Plant Viruses

Plant viruses, like other viruses, contain a core of either DNA or RNA. You have already learned about one of these, the tobacco mosaic virus. As plant cells have a cell wall to protect their cells, these viruses do not use receptor-mediated endocytosis to enter host cells as is seen with animal viruses. For many plant viruses to be transferred from plant to plant, damage to some of the plants' cells must occur to allow the virus to enter a new host. This damage is often caused by weather, insects, animals, fire, or human activities like farming or landscaping. Additionally, plant offspring may

inherit viral diseases from parent plants. Plant viruses can be transmitted by a variety of vectors, through contact with an infected plant’s sap, by living organisms such as insects and nematodes, and through pollen. When plants viruses are transferred between different plants, this is known as **horizontal transmission**, and when they are inherited from a parent, this is called **vertical transmission**.

Symptoms of viral diseases vary according to the virus and its host ([link]). One common symptom is **hyperplasia**, the abnormal proliferation of cells that causes the appearance of plant tumors known as **galls**. Other viruses induce **hypoplasia**, or decreased cell growth, in the leaves of plants, causing thin, yellow areas to appear. Still other viruses affect the plant by directly killing plant cells, a process known as **cell necrosis**. Other symptoms of plant viruses include malformed leaves, black streaks on the stems of the plants, altered growth of stems, leaves, or fruits, and ring spots, which are circular or linear areas of discoloration found in a leaf.

Some Common Symptoms of Plant Viral Diseases	
Symptom	**Appears as**
Hyperplasia	Galls (tumors)
Hypoplasia	Thinned, yellow splotches on leaves
Cell necrosis	Dead, blackened stems, leaves, or fruit
Abnormal growth patterns	Malformed stems, leaves, or fruit
Discoloration	Yellow, red, or black lines, or rings in stems, leaves, or fruit

Plant viruses can seriously disrupt crop growth and development, significantly affecting our food supply. They are responsible for poor crop quality and quantity globally, and can bring about huge economic losses annually. Others viruses may damage plants used in landscaping. Some viruses that infect agricultural food plants include the name of the plant they infect, such as tomato spotted wilt virus, bean common mosaic virus, and cucumber mosaic virus. In plants used for landscaping, two of the most common viruses are peony ring spot and rose mosaic virus. There are far too many plant viruses to discuss each in detail, but symptoms of bean common mosaic virus result in lowered bean production and stunted, unproductive plants. In the ornamental rose, the rose mosaic disease causes wavy yellow lines and colored splotches on the leaves of the plant.

Section Summary

Viral replication within a living cell always produces changes in the cell, sometimes resulting in cell death and sometimes slowly killing the infected cells. There are six basic stages in the virus replication cycle: attachment, penetration, uncoating, replication, assembly, and release. A viral infection may be productive, resulting in new virions, or nonproductive, which means that the virus remains inside the cell without producing new virions. Bacteriophages are viruses that infect bacteria. They have two different modes of replication: the lytic cycle, where the virus replicates and bursts out of the bacteria, and the lysogenic cycle, which involves the incorporation of the viral genome into the bacterial host genome. Animal viruses cause a variety of infections, with some causing chronic symptoms (hepatitis C), some intermittent symptoms (latent viruses such a herpes simplex virus 1), and others that cause very few symptoms, if any (human herpesviruses 6 and 7). Oncogenic viruses in animals have the ability to cause cancer by interfering with the regulation of the host cell cycle. Viruses of plants are responsible for significant economic damage in both agriculture and plants used for ornamentation.

Art Connections

Exercise:

Problem:

[link] Influenza virus is packaged in a viral envelope that fuses with the plasma membrane. This way, the virus can exit the host cell without killing it. What advantage does the virus gain by keeping the host cell alive?

Solution:

[link] The host cell can continue to make new virus particles.

Exercise:

Problem: [link] Which of the following statements is false?

a. In the lytic cycle, new phage are produced and released into the environment.
b. In the lysogenic cycle, phage DNA is incorporated into the host genome.
c. An environmental stressor can cause the phage to initiate the lysogenic cycle.
d. Cell lysis only occurs in the lytic cycle.

Solution:

[link] C

Review Questions

Exercise:

Problem: Which statement is *not* true of viral replication?

a. A lysogenic cycle kills the host cell.
b. There are six basic steps in the viral replication cycle.
c. Viral replication does not affect host cell function.

d. Newly released virions can infect adjacent cells.

Solution:

D

Exercise:

Problem: Which statement is true of viral replication?

a. In the process of apoptosis, the cell survives.
b. During attachment, the virus attaches at specific sites on the cell surface.
c. The viral capsid helps the host cell produce more copies of the viral genome.
d. mRNA works outside of the host cell to produce enzymes and proteins.

Solution:

B

Exercise:

Problem: Which statement is true of reverse transcriptase?

a. It is a nucleic acid.
b. It infects cells.
c. It transcribes RNA to make DNA.
d. It is a lipid.

Solution:

C

Exercise:

Problem: Oncogenic virus cores can be_______.

a. RNA
b. DNA
c. neither RNA nor DNA
d. either RNA or DNA

Solution:

D

Exercise:

Problem: Which is true of DNA viruses?

a. They use the host cell’s machinery to produce new copies of their genome.
b. They all have envelopes.
c. They are the only kind of viruses that can cause cancer.
d. They are not important plant pathogens.

Solution:

A

Exercise:

Problem:A bacteriophage can infect ________.

a. the lungs
b. viruses
c. prions
d. bacteria

Solution:

D

Free Response

Exercise:

Problem: Why can’t dogs catch the measles?

Solution:

The virus can’t attach to dog cells, because dog cells do not express the receptors for the virus and/or there is no cell within the dog that is permissive for viral replication.

Exercise:

Problem:

One of the first and most important targets for drugs to fight infection with HIV (a retrovirus) is the reverse transcriptase enzyme. Why?

Solution:

Reverse transcriptase is needed to make more HIV-1 viruses, so targeting the reverse transcriptase enzyme may be a way to inhibit the replication of the virus. Importantly, by targeting reverse transcriptase, we do little harm to the host cell, since host cells do not make reverse transcriptase. Thus, we can specifically attack the virus and not the host cell when we use reverse transcriptase inhibitors.

Exercise:

Problem:

In this section, you were introduced to different types of viruses and viral diseases. Briefly discuss the most interesting or surprising thing you learned about viruses.

Solution:

Answer is open and will vary.

Exercise:

Problem:

Although plant viruses cannot infect humans, what are some of the ways in which they affect humans?

Solution:

Plant viruses infect crops, causing crop damage and failure, and considerable economic losses.

Glossary

acute disease
: disease where the symptoms rise and fall within a short period of time

asymptomatic disease
: disease where there are no symptoms and the individual is unaware of being infected unless lab tests are performed

AZT
: anti-HIV drug that inhibits the viral enzyme reverse transcriptase

bacteriophage
: virus that infects bacteria

budding

method of exit from the cell used in certain animal viruses, where virions leave the cell individually by capturing a piece of the host plasma membrane

cell necrosis
: cell death

chronic infection
: describes when the virus persists in the body for a long period of time

cytopathic
: causing cell damage

fusion
: method of entry by some enveloped viruses, where the viral envelope fuses with the plasma membrane of the host cell

gall
: appearance of a plant tumor

horizontal transmission
: transmission of a disease between unrelated individuals

hyperplasia
: abnormally high cell growth and division

hypoplasia
: abnormally low cell growth and division

intermittent symptom
: symptom that occurs periodically

latency
: virus that remains in the body for a long period of time but only causes intermittent symptoms

lysis
: bursting of a cell

lytic cycle
: type of virus replication in which virions are released through lysis, or bursting, of the cell

lysogenic cycle
: type of virus replication in which the viral genome is incorporated into the genome of the host cell

oncogenic virus
: virus that has the ability to cause cancer

permissive
: cell type that is able to support productive replication of a virus

productive
: viral infection that leads to the production of new virions

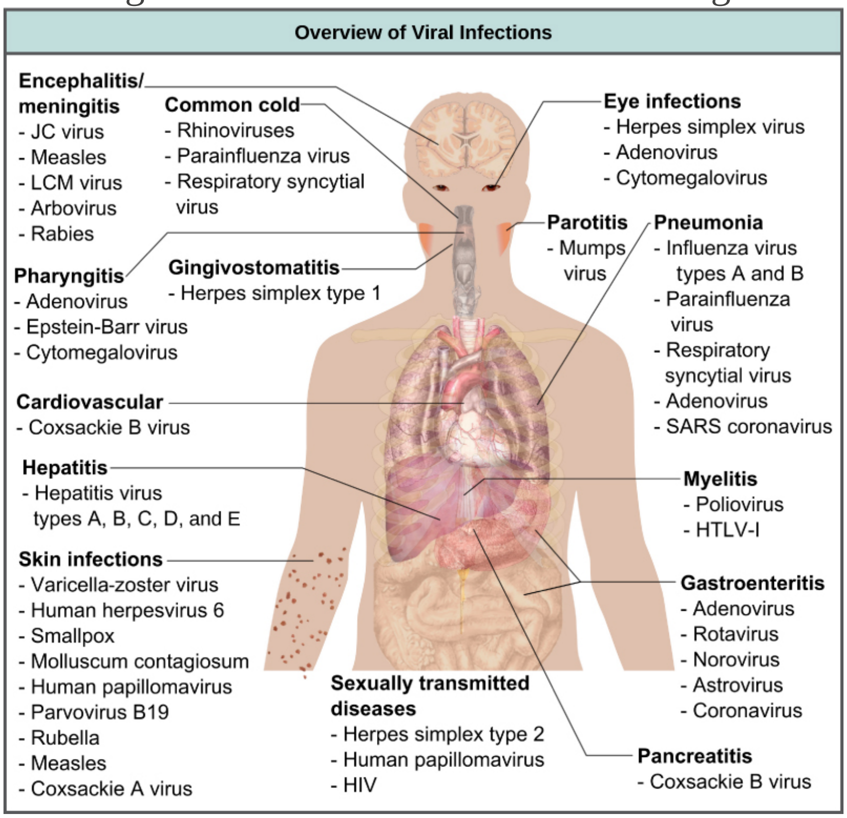
prophage
: phage DNA that is incorporated into the host cell genome

vertical transmission
: transmission of disease from parent to offspring

Prevention and Treatment of Viral Infections
By the end of this section, you will be able to:

- Identify major viral illnesses that affect humans
- Compare vaccinations and anti-viral drugs as medical approaches to viruses

Viruses cause a variety of diseases in animals, including humans, ranging from the common cold to potentially fatal illnesses like meningitis ([link]). These diseases can be treated by antiviral drugs or by vaccines, but some viruses, such as HIV, are capable of both avoiding the immune response and mutating to become resistant to antiviral drugs.



Viruses can cause dozens of ailments in humans, ranging from mild illnesses to serious diseases. (credit: modification of work by Mikael Häggström)

Vaccines for Prevention

While we do have limited numbers of effective antiviral drugs, such as those used to treat HIV and influenza, the primary method of controlling viral disease is by vaccination, which is intended to prevent outbreaks by building immunity to a virus or virus family ([link]). **Vaccines** may be prepared using live viruses, killed viruses, or molecular subunits of the virus. The killed viral vaccines and subunit viruses are both incapable of causing disease.

Live viral vaccines are designed in the laboratory to cause few symptoms in recipients while giving them protective immunity against future infections. Polio was one disease that represented a milestone in the use of vaccines. Mass immunization campaigns in the 1950s (killed vaccine) and 1960s (live vaccine) significantly reduced the incidence of the disease, which caused muscle paralysis in children and generated a great amount of fear in the general population when regional epidemics occurred. The success of the polio vaccine paved the way for the routine dispensation of childhood vaccines against measles, mumps, rubella, chickenpox, and other diseases.

The danger of using live vaccines, which are usually more effective than killed vaccines, is the low but significant danger that these viruses will revert to their disease-causing form by **back mutations**. Live vaccines are usually made by **attenuating** (weakening) the “wild-type” (disease-causing) virus by growing it in the laboratory in tissues or at temperatures different from what the virus is accustomed to in the host. Adaptations to these new cells or temperatures induce mutations in the genomes of the virus, allowing it to grow better in the laboratory while inhibiting its ability to cause disease when reintroduced into conditions found in the host. These attenuated viruses thus still cause infection, but they do not grow very well, allowing the immune response to develop in time to prevent major disease. Back mutations occur when the vaccine undergoes mutations in the host such that it readapts to the host and can again cause disease, which can then be spread to other humans in an epidemic. This type of scenario happened as recently as 2007 in Nigeria where mutations in a polio vaccine led to an epidemic of polio in that country.

Some vaccines are in continuous development because certain viruses, such as influenza and HIV, have a high mutation rate compared to other viruses and normal host cells. With influenza, mutations in the surface molecules of the virus help the organism evade the protective immunity that may have been obtained in a previous influenza season, making it necessary for individuals to get vaccinated every year. Other viruses, such as those that cause the childhood diseases measles, mumps, and rubella, mutate so infrequently that the same vaccine is used year after year.

Vaccinations are designed to boost immunity to a virus to prevent infection. (credit: USACE Europe District)

Note:
Link to Learning

Watch this NOVA video to learn how microbiologists are attempting to replicate the deadly 1918 Spanish influenza virus so they can understand more about virology.

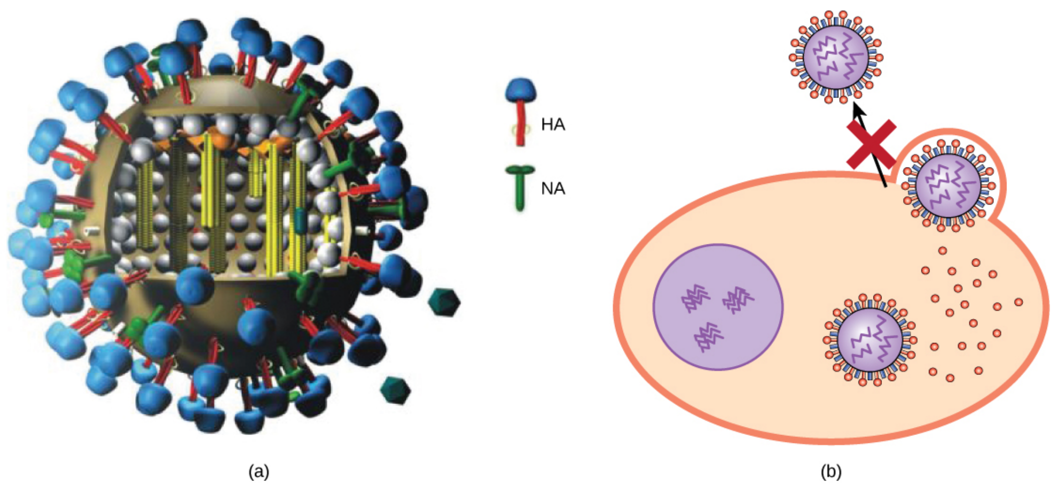
Vaccines and Anti-viral Drugs for Treatment

In some cases, vaccines can be used to treat an active viral infection. The concept behind this is that by giving the vaccine, immunity is boosted without adding more disease-causing virus. In the case of rabies, a fatal neurological disease transmitted via the saliva of rabies virus-infected animals, the progression of the disease from the time of the animal bite to the time it enters the central nervous system may be 2 weeks or longer. This is enough time to vaccinate an individual who suspects that they have been bitten by a rabid animal, and their boosted immune response is sufficient to prevent the virus from entering nervous tissue. Thus, the potentially fatal neurological consequences of the disease are averted, and the individual only has to recover from the infected bite. This approach is also being used for the treatment of Ebola, one of the fastest and most deadly viruses on earth. Transmitted by bats and great apes, this disease can cause death in 70–90 percent of infected humans within 2 weeks. Using newly developed vaccines that boost the immune response in this way, there is hope that affected individuals will be better able to control the virus, potentially saving a greater percentage of infected persons from a rapid and very painful death.

Another way of treating viral infections is the use of antiviral drugs. These drugs often have limited success in curing viral disease, but in many cases, they have been used to control and reduce symptoms for a wide variety of viral diseases. For most viruses, these drugs can inhibit the virus by

blocking the actions of one or more of its proteins. It is important that the targeted proteins be encoded by viral genes and that these molecules are not present in a healthy host cell. In this way, viral growth is inhibited without damaging the host. There are large numbers of antiviral drugs available to treat infections, some specific for a particular virus and others that can affect multiple viruses.

Antivirals have been developed to treat genital herpes (herpes simplex II) and influenza. For genital herpes, drugs such as acyclovir can reduce the number and duration of episodes of active viral disease, during which patients develop viral lesions in their skin cells. As the virus remains latent in nervous tissue of the body for life, this drug is not curative but can make the symptoms of the disease more manageable. For influenza, drugs like Tamiflu (oseltamivir) ([link]) can reduce the duration of "flu" symptoms by 1 or 2 days, but the drug does not prevent symptoms entirely. Tamiflu works by inhibiting an enzyme (viral neuraminidase) that allows new virions to leave their infected cells. Thus, Tamiflu inhibits the spread of virus from infected to uninfected cells. Other antiviral drugs, such as Ribavirin, have been used to treat a variety of viral infections, although its mechanism of action against certain viruses remains unclear.

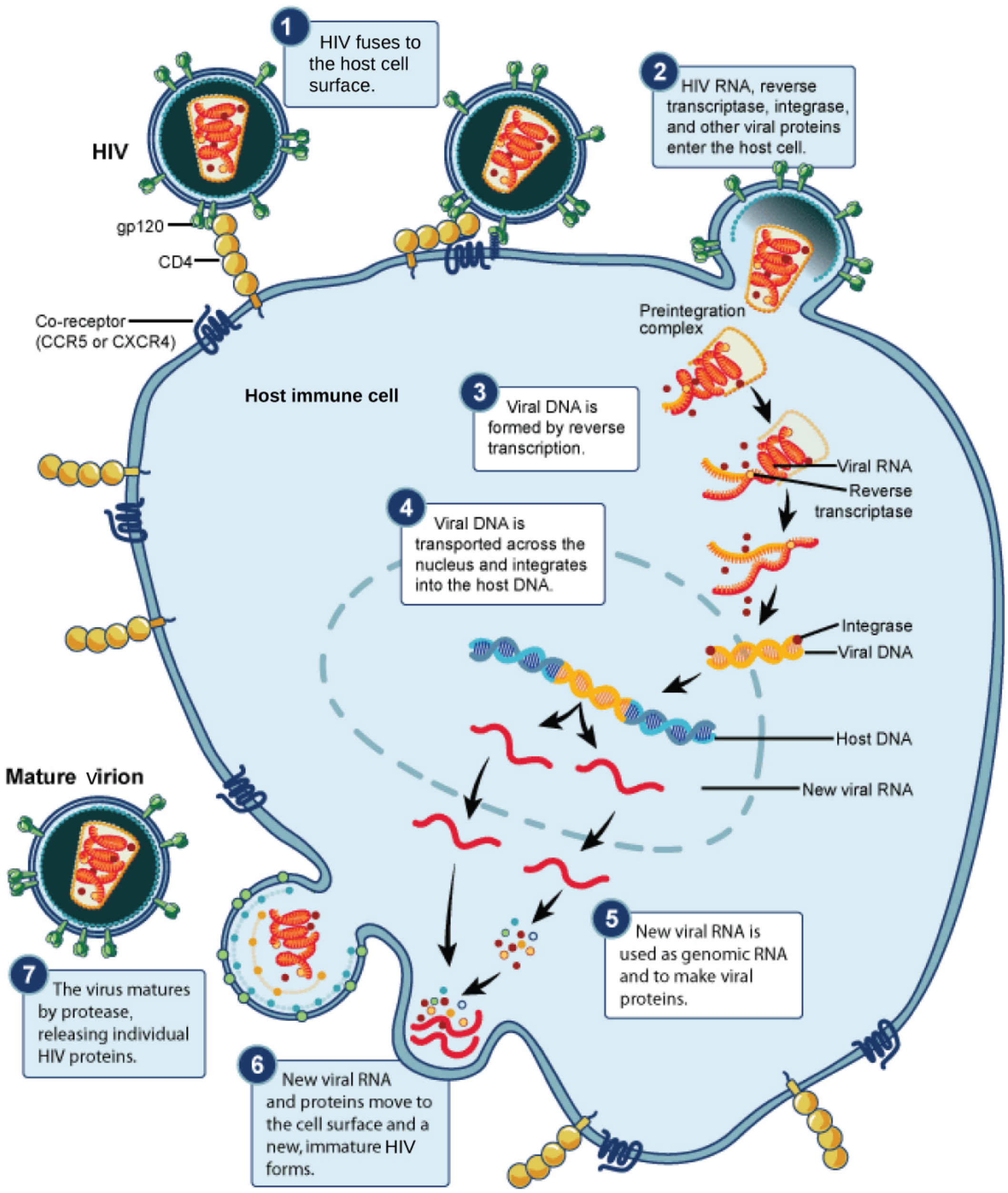


(a) Tamiflu inhibits a viral enzyme called neuraminidase (NA) found in the influenza viral envelope. (b) Neuraminidase cleaves the connection between viral

hemagglutinin (HA), also found in the viral envelope, and glycoproteins on the host cell surface. Inhibition of neuraminidase prevents the virus from detaching from the host cell, thereby blocking further infection. (credit a: modification of work by M. Eickmann)

By far, the most successful use of antivirals has been in the treatment of the retrovirus HIV, which causes a disease that, if untreated, is usually fatal within 10–12 years after infection. Anti-HIV drugs have been able to control viral replication to the point that individuals receiving these drugs survive for a significantly longer time than the untreated.

Anti-HIV drugs inhibit viral replication at many different phases of the HIV replicative cycle ([link]). Drugs have been developed that inhibit the fusion of the HIV viral envelope with the plasma membrane of the host cell (fusion inhibitors), the conversion of its RNA genome into double-stranded DNA (reverse transcriptase inhibitors), the integration of the viral DNA into the host genome (integrase inhibitors), and the processing of viral proteins (protease inhibitors).



HIV, an enveloped, icosahedral virus, attaches to the CD4 receptor of an immune cell and fuses with the cell membrane. Viral contents are released into the cell, where viral enzymes convert the single-stranded RNA genome into DNA and incorporate it into the host genome. (credit: NIAID, NIH)

When any of these drugs are used individually, the high mutation rate of the virus allows it to easily and rapidly develop resistance to the drug, limiting the drug’s effectiveness. The breakthrough in the treatment of HIV was the development of HAART, highly active anti-retroviral therapy, which involves a mixture of different drugs, sometimes called a drug “cocktail.” By attacking the virus at different stages of its replicative cycle, it is much more difficult for the virus to develop resistance to multiple drugs at the same time. Still, even with the use of combination HAART therapy, there is concern that, over time, the virus will develop resistance to this therapy. Thus, new anti-HIV drugs are constantly being developed with the hope of continuing the battle against this highly fatal virus.

Note:
Everyday Connection
Applied Virology
The study of viruses has led to the development of a variety of new ways to treat non-viral diseases. Viruses have been used in **gene therapy**. Gene therapy is used to treat genetic diseases such as severe combined immunodeficiency (SCID), a heritable, recessive disease in which children are born with severely compromised immune systems. One common type of SCID is due to the lack of an enzyme, adenosine deaminase (ADA), which breaks down purine bases. To treat this disease by gene therapy, bone marrow cells are taken from a SCID patient and the ADA gene is inserted. This is where viruses come in, and their use relies on their ability to penetrate living cells and bring genes in with them. Viruses such as adenovirus, an upper respiratory human virus, are modified by the addition of the ADA gene, and the virus then transports this gene into the cell. The modified cells, now capable of making ADA, are then given back to the patients in the hope of curing them. Gene therapy using viruses as carrier of genes (viral vectors), although still experimental, holds promise for the treatment of many genetic diseases. Still, many technological problems need to be solved for this approach to be a viable method for treating genetic disease.
Another medical use for viruses relies on their specificity and ability to kill the cells they infect. **Oncolytic viruses** are engineered in the laboratory

specifically to attack and kill cancer cells. A genetically modified adenovirus known as H101 has been used since 2005 in clinical trials in China to treat head and neck cancers. The results have been promising, with a greater short-term response rate to the combination of chemotherapy and viral therapy than to chemotherapy treatment alone. This ongoing research may herald the beginning of a new age of cancer therapy, where viruses are engineered to find and specifically kill cancer cells, regardless of where in the body they may have spread.

A third use of viruses in medicine relies on their specificity and involves using bacteriophages in the treatment of bacterial infections. Bacterial diseases have been treated with antibiotics since the 1940s. However, over time, many bacteria have developed resistance to antibiotics. A good example is methicillin-resistant *Staphylococcus aureus* (MRSA, pronounced “mersa”), an infection commonly acquired in hospitals. This bacterium is resistant to a variety of antibiotics, making it difficult to treat. The use of bacteriophages specific for such bacteria would bypass their resistance to antibiotics and specifically kill them. Although **phage therapy** is in use in the Republic of Georgia to treat antibiotic-resistant bacteria, its use to treat human diseases has not been approved in most countries. However, the safety of the treatment was confirmed in the United States when the U.S. Food and Drug Administration approved spraying meats with bacteriophages to destroy the food pathogen *Listeria*. As more and more antibiotic-resistant strains of bacteria evolve, the use of bacteriophages might be a potential solution to the problem, and the development of phage therapy is of much interest to researchers worldwide.

Section Summary

Viruses cause a variety of diseases in humans. Many of these diseases can be prevented by the use of viral vaccines, which stimulate protective immunity against the virus without causing major disease. Viral vaccines may also be used in active viral infections, boosting the ability of the immune system to control or destroy the virus. A series of antiviral drugs that target enzymes and other protein products of viral genes have been

developed and used with mixed success. Combinations of anti-HIV drugs have been used to effectively control the virus, extending the lifespans of infected individuals. Viruses have many uses in medicines, such as in the treatment of genetic disorders, cancer, and bacterial infections.

Review Questions

Exercise:

Problem:

Which of the following is NOT used to treat active viral disease?

a. vaccines
b. antiviral drugs
c. antibiotics
d. phage therapy

Solution:

C

Exercise:

Problem: Vaccines________.

a. are similar to viroids
b. are only needed once
c. kill viruses
d. stimulate an immune response

Solution:

D

Free Response

Exercise:

Problem:

Why is immunization after being bitten by a rabid animal so effective and why aren’t people vaccinated for rabies like dogs and cats are?

Solution:

Rabies vaccine works after a bite because it takes week for the virus to travel from the site of the bite to the central nervous system, where the most severe symptoms of the disease occur. Adults are not routinely vaccinated for rabies for two reasons: first, because the routine vaccination of domestic animals makes it unlikely that humans will contract rabies from an animal bite; second, if one is bitten by a wild animal or a domestic animal that one cannot confirm has been immunized, there is still time to give the vaccine and avoid the often fatal consequences of the disease.

Glossary

attenuation
: weakening of a virus during vaccine development

back mutation
: when a live virus vaccine reverts back to it disease-causing phenotype

gene therapy
: treatment of genetic disease by adding genes, using viruses to carry the new genes inside the cell

oncolytic virus
: virus engineered to specifically infect and kill cancer cells

phage therapy
: treatment of bacterial diseases using bacteriophages specific to a particular bacterium

vaccine

weakened solution of virus components, viruses, or other agents that produce an immune response

Other Acellular Entities: Prions and Viroids
By the end of this section, you will be able to:

- Describe prions and their basic properties
- Define viroids and their targets of infection

Prions and viroids are **pathogens** (agents with the ability to cause disease) that have simpler structures than viruses but, in the case of prions, still can produce deadly diseases.

Prions

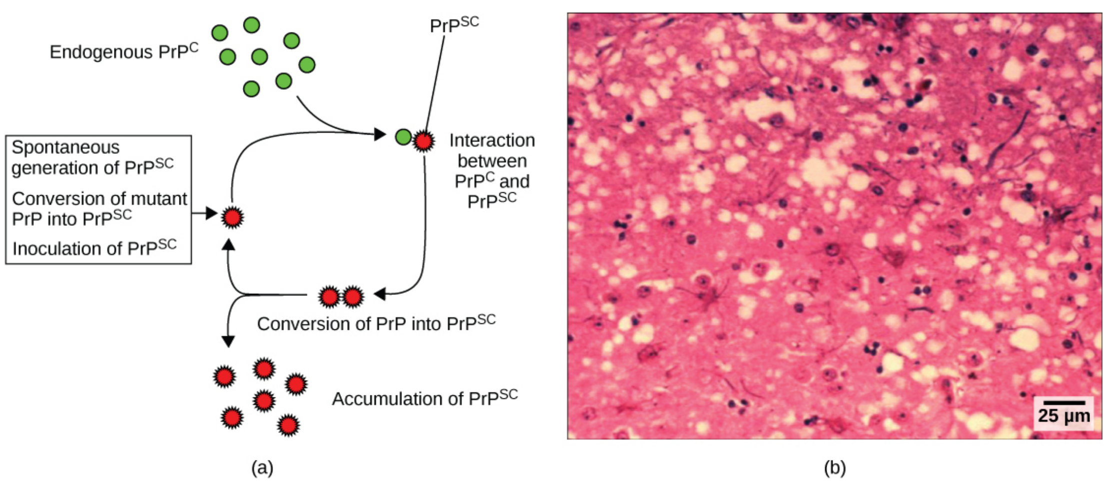
Prions, so-called because they are proteinaceous, are infectious particles—smaller than viruses—that contain no nucleic acids (neither DNA nor RNA). Historically, the idea of an infectious agent that did not use nucleic acids was considered impossible, but pioneering work by Nobel Prize-winning biologist Stanley Prusiner has convinced the majority of biologists that such agents do indeed exist.

Fatal neurodegenerative diseases, such as kuru in humans and bovine spongiform encephalopathy (BSE) in cattle (commonly known as "mad cow disease") were shown to be transmitted by prions. The disease was spread by the consumption of meat, nervous tissue, or internal organs between members of the same species. Kuru, native to humans in Papua New Guinea, was spread from human to human via ritualistic cannibalism. BSE, originally detected in the United Kingdom, was spread between cattle by the practice of including cattle nervous tissue in feed for other cattle. Individuals with kuru and BSE show symptoms of loss of motor control and unusual behaviors, such as uncontrolled bursts of laughter with kuru, followed by death. Kuru was controlled by inducing the population to abandon its ritualistic cannibalism.

On the other hand, BSE was initially thought to only affect cattle. Cattle dying of the disease were shown to have developed lesions or "holes" in the brain, causing the brain tissue to resemble a sponge. Later on in the outbreak, however, it was shown that a similar encephalopathy in humans known as variant Creutzfeldt-Jakob disease (CJD) could be acquired from

eating beef from animals with BSE, sparking bans by various countries on the importation of British beef and causing considerable economic damage to the British beef industry ([link]). BSE still exists in various areas, and although a rare disease, individuals that acquire CJD are difficult to treat. The disease can be spread from human to human by blood, so many countries have banned blood donation from regions associated with BSE.

The cause of spongiform encephalopathies, such as kuru and BSE, is an infectious structural variant of a normal cellular protein called PrP (prion protein). It is this variant that constitutes the prion particle. PrP exists in two forms, **PrP^c**, the normal form of the protein, and **PrP^{sc}**, the infectious form. Once introduced into the body, the PrP^{sc} contained within the prion binds to PrP^c and converts it to PrP^{sc}. This leads to an exponential increase of the PrP^{sc} protein, which aggregates. PrP^{sc} is folded abnormally, and the resulting conformation (shape) is directly responsible for the lesions seen in the brains of infected cattle. Thus, although not without some detractors among scientists, the prion seems likely to be an entirely new form of infectious agent, the first one found whose transmission is not reliant upon genes made of DNA or RNA.



(a) Endogenous normal prion protein (PrP^c) is converted into the disease-causing form (PrP^{sc}) when it encounters this variant form of the protein. PrP^{sc} may arise spontaneously in brain tissue, especially if a mutant form of the protein is present, or it

may occur via the spread of misfolded prions consumed in food into brain tissue. (b) This prion-infected brain tissue, visualized using light microscopy, shows the vacuoles that give it a spongy texture, typical of transmissible spongiform encephalopathies. (credit b: modification of work by Dr. Al Jenny, USDA APHIS; scale-bar data from Matt Russell)

Viroids

Viroids are plant pathogens: small, single-stranded, circular RNA particles that are much simpler than a virus. They do not have a capsid or outer envelope, but like viruses can reproduce only within a host cell. Viroids do not, however, manufacture any proteins, and they only produce a single, specific RNA molecule. Human diseases caused by viroids have yet to be identified.

Viroids are known to infect plants ([link]) and are responsible for crop failures and the loss of millions of dollars in agricultural revenue each year. Some of the plants they infect include potatoes, cucumbers, tomatoes, chrysanthemums, avocados, and coconut palms.

These potatoes have been infected by the potato spindle tuber viroid (PSTV), which is typically spread when infected knives are used to cut healthy potatoes, which are then planted. (credit: Pamela Roberts, University of Florida Institute of Food and Agricultural Sciences, USDA ARS)

Note:
Career Connection
Virologist
Virology is the study of viruses, and a virologist is an individual trained in this discipline. Training in virology can lead to many different career paths. Virologists are actively involved in academic research and teaching in colleges and medical schools. Some virologists treat patients or are involved in the generation and production of vaccines. They might participate in epidemiologic studies ([link]) or become science writers, to name just a few possible careers.

This virologist is engaged in fieldwork, sampling eggs from this nest for avian influenza. (credit: Don Becker, USGS EROS, U.S. Fish and Wildlife Service)

If you think you may be interested in a career in virology, find a mentor in the field. Many large medical centers have departments of virology, and smaller hospitals usually have virology labs within their microbiology departments. Volunteer in a virology lab for a semester or work in one over the summer. Discussing the profession and getting a first-hand look at the work will help you decide whether a career in virology is right for you. The American Society of Virology’s website is a good resource for information regarding training and careers in virology.

Section Summary

Prions are infectious agents that consist of protein, but no DNA or RNA, and seem to produce their deadly effects by duplicating their shapes and accumulating in tissues. They are thought to contribute to several progressive brain disorders, including mad cow disease and Creutzfeldt-Jakob disease. Viroids are single-stranded RNA pathogens that infect plants. Their presence can have a severe impact on the agriculture industry.

Review Questions

Exercise:

Problem: Which of the following is not associated with prions?

a. replicating shapes
b. mad cow disease
c. DNA
d. toxic proteins

Solution:

C

Exercise:

Problem: Which statement is true of viroids?

a. They are single-stranded RNA particles.
b. They reproduce only outside of the cell.
c. They produce proteins.
d. They affect both plants and animals.

Solution:

A

Free Response

Exercise:

Problem:

Prions are responsible for variant Creutzfeldt-Jakob Disease, which has resulted in over 100 human deaths in Great Britain during the last 10 years. How do humans obtain this disease?

Solution:

This prion-based disease is transmitted through human consumption of infected meat.

Exercise:

Problem: How are viroids like viruses?

Solution:

They both replicate in a cell, and they both contain nucleic acid.

Glossary

pathogen
: agent with the ability to cause disease

prion
: infectious particle that consists of proteins that replicate without DNA or RNA

PrP^{c}
: normal prion protein

PrP^{sc}
: infectious form of a prion protein

viroid

plant pathogen that produces only a single, specific RNA

Prokaryotic Diversity
By the end of this section, you will be able to:

- Describe the evolutionary history of prokaryotes
- Discuss the distinguishing features of extremophiles
- Explain why it is difficult to culture prokaryotes

Prokaryotes are ubiquitous. They cover every imaginable surface where there is sufficient moisture, and they live on and inside of other living things. In the typical human body, prokaryotic cells outnumber human body cells by about ten to one. They comprise the majority of living things in all ecosystems. Some prokaryotes thrive in environments that are inhospitable for most living things. Prokaryotes recycle **nutrients**—essential substances (such as carbon and nitrogen)—and they drive the evolution of new ecosystems, some of which are natural and others man-made. Prokaryotes have been on Earth since long before multicellular life appeared.

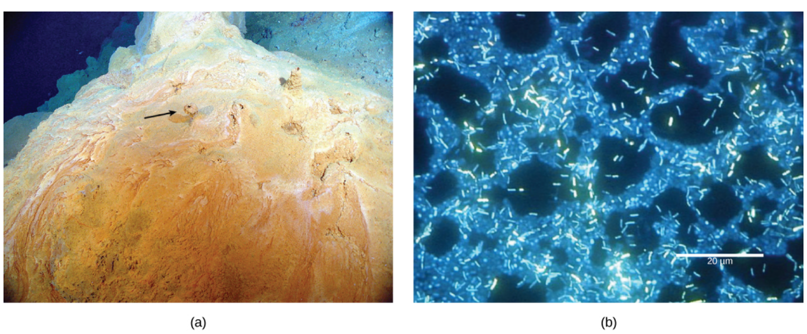
Prokaryotes, the First Inhabitants of Earth

When and where did life begin? What were the conditions on Earth when life began? Prokaryotes were the first forms of life on Earth, and they existed for billions of years before plants and animals appeared. The Earth and its moon are thought to be about 4.54 billion years old. This estimate is based on evidence from radiometric dating of meteorite material together with other substrate material from Earth and the moon. Early Earth had a very different atmosphere (contained less molecular oxygen) than it does today and was subjected to strong radiation; thus, the first organisms would have flourished where they were more protected, such as in ocean depths or beneath the surface of the Earth. At this time too, strong volcanic activity was common on Earth, so it is likely that these first organisms—the first prokaryotes—were adapted to very high temperatures. Early Earth was prone to geological upheaval and volcanic eruption, and was subject to bombardment by mutagenic radiation from the sun. The first organisms were prokaryotes that could withstand these harsh conditions.

Microbial Mats

Microbial mats or large biofilms may represent the earliest forms of life on Earth; there is fossil evidence of their presence starting about 3.5 billion years ago. A **microbial mat** is a multi-layered sheet of prokaryotes ([link]) that includes mostly bacteria, but also archaea. Microbial mats are a few centimeters thick, and they typically grow where different types of materials interface, mostly on moist surfaces. The various types of prokaryotes that comprise them carry out different metabolic pathways, and that is the reason for their various colors. Prokaryotes in a microbial mat are held together by a glue-like sticky substance that they secrete called extracellular matrix.

The first microbial mats likely obtained their energy from chemicals found near hydrothermal vents. A **hydrothermal vent** is a breakage or fissure in the Earth’s surface that releases geothermally heated water. With the evolution of photosynthesis about 3 billion years ago, some prokaryotes in microbial mats came to use a more widely available energy source—sunlight—whereas others were still dependent on chemicals from hydrothermal vents for energy and food.



This (a) microbial mat, about one meter in diameter, grows over a hydrothermal vent in the Pacific Ocean in a region known as the “Pacific Ring of Fire.” The mat helps retain microbial nutrients. Chimneys such as the one indicated by the arrow allow gases to escape. (b) In this micrograph, bacteria are

visualized using fluorescence microscopy. (credit a: modification of work by Dr. Bob Embley, NOAA PMEL, Chief Scientist; credit b: modification of work by Ricardo Murga, Rodney Donlan, CDC; scale-bar data from Matt Russell)

Stromatolites

Fossilized microbial mats represent the earliest record of life on Earth. A **stromatolite** is a sedimentary structure formed when minerals are precipitated out of water by prokaryotes in a microbial mat ([link]). Stromatolites form layered rocks made of carbonate or silicate. Although most stromatolites are artifacts from the past, there are places on Earth where stromatolites are still forming. For example, growing stromatolites have been found in the Anza-Borrego Desert State Park in San Diego County, California.

(a) These living stromatolites are located in Shark Bay, Australia. (b) These fossilized stromatolites, found in Glacier National Park, Montana, are nearly 1.5 billion years old. (credit a: Robert Young; credit b: P. Carrara, NPS)

The Ancient Atmosphere

Evidence indicates that during the first two billion years of Earth's existence, the atmosphere was **anoxic**, meaning that there was no molecular oxygen. Therefore, only those organisms that can grow without oxygen—**anaerobic** organisms—were able to live. Autotrophic organisms that convert solar energy into chemical energy are called **phototrophs**, and they appeared within one billion years of the formation of Earth. Then, **cyanobacteria**, also known as blue-green algae, evolved from these simple phototrophs one billion years later. Cyanobacteria ([link]) began the oxygenation of the atmosphere. Increased atmospheric oxygen allowed the development of more efficient O_2-utilizing catabolic pathways. It also opened up the land to increased colonization, because some O_2 is converted into O_3 (ozone) and ozone effectively absorbs the ultraviolet light that would otherwise cause lethal mutations in DNA. Ultimately, the increase in O_2 concentrations allowed the evolution of other life forms.

This hot spring in Yellowstone National Park flows toward the foreground. Cyanobacteria in the spring are green, and as

water flows down the gradient, the intensity of the color increases as cell density increases. The water is cooler at the edges of the stream than in the center, causing the edges to appear greener. (credit: Graciela Brelles-Mariño)

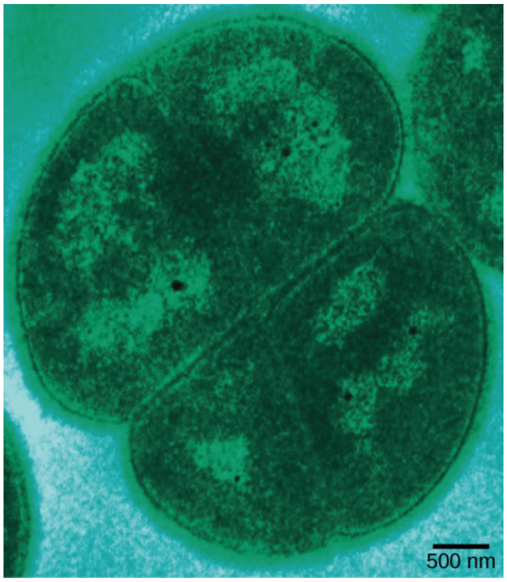
Microbes Are Adaptable: Life in Moderate and Extreme Environments

Some organisms have developed strategies that allow them to survive harsh conditions. Prokaryotes thrive in a vast array of environments: Some grow in conditions that would seem very normal to us, whereas others are able to thrive and grow under conditions that would kill a plant or animal. Almost all prokaryotes have a cell wall, a protective structure that allows them to survive in both hyper- and hypo-osmotic conditions. Some soil bacteria are able to form endospores that resist heat and drought, thereby allowing the organism to survive until favorable conditions recur. These adaptations, along with others, allow bacteria to be the most abundant life form in all terrestrial and aquatic ecosystems.

Other bacteria and archaea are adapted to grow under extreme conditions and are called **extremophiles**, meaning “lovers of extremes.” Extremophiles have been found in all kinds of environments: the depth of the oceans, hot springs, the Artic and the Antarctic, in very dry places, deep inside Earth, in harsh chemical environments, and in high radiation environments ([link]), just to mention a few. These organisms give us a better understanding of prokaryotic diversity and open up the possibility of finding new prokaryotic species that may lead to the discovery of new therapeutic drugs or have industrial applications. Because they have specialized adaptations that allow them to live in extreme conditions, many extremophiles cannot survive in moderate environments. There are many different groups of extremophiles: They are identified based on the

conditions in which they grow best, and several habitats are extreme in multiple ways. For example, a soda lake is both salty and alkaline, so organisms that live in a soda lake must be both alkaliphiles and halophiles ([link]). Other extremophiles, like **radioresistant** organisms, do not prefer an extreme environment (in this case, one with high levels of radiation), but have adapted to survive in it ([link]).

Extremophiles and Their Preferred Conditions	
Extremophile Type	**Conditions for Optimal Growth**
Acidophiles	pH 3 or below
Alkaliphiles	pH 9 or above
Thermophiles	Temperature 60–80 °C (140–176 °F)
Hyperthermophiles	Temperature 80–122 °C (176–250 °F)
Psychrophiles	Temperature of -15-10 °C (5-50 °F) or lower
Halophiles	Salt concentration of at least 0.2 M
Osmophiles	High sugar concentration



Deinococcus radiodurans, visualized in this false color transmission electron micrograph, is a prokaryote that can tolerate very high doses of ionizing radiation. It has developed DNA repair mechanisms that allow it to reconstruct its chromosome even if it has been broken into hundreds of pieces by radiation or heat. (credit: modification of work by Michael Daly; scale-bar data from Matt Russell)

Prokaryotes in the Dead Sea

One example of a very harsh environment is the Dead Sea, a hypersaline basin that is located between Jordan and Israel. Hypersaline environments are essentially concentrated seawater. In the Dead Sea, the sodium concentration is 10 times higher than that of seawater, and the water contains high levels of magnesium (about 40 times higher than in seawater) that would be toxic to most living things. Iron, calcium, and magnesium, elements that form divalent ions (Fe^{2+}, Ca^{2+}, and Mg^{2+}), produce what is commonly referred to as “hard” water. Taken together, the high concentration of divalent cations, the acidic pH (6.0), and the intense solar radiation flux make the Dead Sea a unique, and uniquely hostile, ecosystem[footnote] ([link]).

Bodaker, I, Itai, S, Suzuki, MT, Feingersch, R, Rosenberg, M, Maguire, ME, Shimshon, B, and others. Comparative community genomics in the Dead Sea: An increasingly extreme environment. *The ISME Journal* 4 (2010): 399–407, doi:10.1038/ismej.2009.141. published online 24 December 2009.

What sort of prokaryotes do we find in the Dead Sea? The extremely salt-tolerant bacterial mats include *Halobacterium*, *Haloferax volcanii* (which is found in other locations, not only the Dead Sea), *Halorubrum sodomense*, and *Halobaculum gomorrense*, and the archaea *Haloarcula marismortui*, among others.

(a) (b)

(a) The Dead Sea is hypersaline. Nevertheless, salt-tolerant bacteria thrive in this sea. (b) These halobacteria cells can form salt-tolerant bacterial

mats. (credit a: Julien Menichini; credit b: NASA; scale-bar data from Matt Russell)

Unculturable Prokaryotes and the Viable-but-Non-Culturable State

Microbiologists typically grow prokaryotes in the laboratory using an appropriate culture medium containing all the nutrients needed by the target organism. The medium can be liquid, broth, or solid. After an incubation time at the right temperature, there should be evidence of microbial growth ([link]). The process of culturing bacteria is complex and is one of the greatest discoveries of modern science. German physician Robert Koch is credited with discovering the techniques for pure culture, including staining and using growth media. His assistant Julius Petri invented the Petri dish whose use persists in today's laboratories. Koch worked primarily with the *Mycobacterium tuberculosis* bacterium that causes tuberculosis and developed postulates to identify disease-causing organisms that continue to be widely used in the medical community. Koch's postulates include that an organism can be identified as the cause of disease when it is present in all infected samples and absent in all healthy samples, and it is able to reproduce the infection after being cultured multiple times. Today, cultures remain a primary diagnostic tool in medicine and other areas of molecular biology.

In these agar plates, the

growth medium is supplemented with red blood cells. Blood agar becomes transparent in the presence of hemolytic *Streptococcus*, which destroys red blood cells and is used to diagnose *Streptococcus* infections. The plate on the left is inoculated with non-hemolytic *Staphylococcus* (large white colonies), and the plate on the right is inoculated with hemolytic *Streptococcus* (tiny clear colonies). If you look closely at the right plate, you can see that the agar surrounding the bacteria has turned clear. (credit: Bill Branson, NCI)

Some prokaryotes, however, cannot grow in a laboratory setting. In fact, over 99 percent of bacteria and archaea are unculturable. For the most part, this is due to a lack of knowledge as to what to feed these organisms and how to grow them; they have special requirements for growth that remain unknown to scientists, such as needing specific micronutrients, pH, temperature, pressure, co-factors, or co-metabolites. Some bacteria cannot be cultured because they are obligate intracellular parasites and cannot be grown outside a host cell.

In other cases, culturable organisms become unculturable under stressful conditions, even though the same organism could be cultured previously. Those organisms that cannot be cultured but are not dead are in a **viable-but-non-culturable** (VBNC) state. The VBNC state occurs when prokaryotes respond to environmental stressors by entering a dormant state

that allows their survival. The criteria for entering into the VBNC state are not completely understood. In a process called **resuscitation**, the prokaryote can go back to “normal” life when environmental conditions improve.

Is the VBNC state an unusual way of living for prokaryotes? In fact, most of the prokaryotes living in the soil or in oceanic waters are non-culturable. It has been said that only a small fraction, perhaps one percent, of prokaryotes can be cultured under laboratory conditions. If these organisms are non-culturable, then how is it known whether they are present and alive? Microbiologists use molecular techniques, such as the polymerase chain reaction (PCR), to amplify selected portions of DNA of prokaryotes, demonstrating their existence. Recall that PCR can make billions of copies of a DNA segment in a process called amplification.

The Ecology of Biofilms

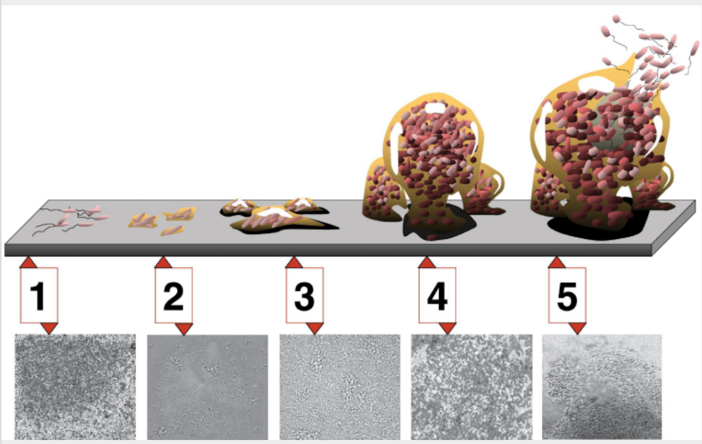
Until a couple of decades ago, microbiologists used to think of prokaryotes as isolated entities living apart. This model, however, does not reflect the true ecology of prokaryotes, most of which prefer to live in communities where they can interact. A **biofilm** is a microbial community ([link]) held together in a gummy-textured matrix that consists primarily of polysaccharides secreted by the organisms, together with some proteins and nucleic acids. Biofilms grow attached to surfaces. Some of the best-studied biofilms are composed of prokaryotes, although fungal biofilms have also been described as well as some composed of a mixture of fungi and bacteria.

Biofilms are present almost everywhere: they can cause the clogging of pipes and readily colonize surfaces in industrial settings. In recent, large-scale outbreaks of bacterial contamination of food, biofilms have played a major role. They also colonize household surfaces, such as kitchen counters, cutting boards, sinks, and toilets, as well as places on the human body, such as the surfaces of our teeth.

Interactions among the organisms that populate a biofilm, together with their protective exopolysaccharidic (EPS) environment, make these

communities more robust than free-living, or planktonic, prokaryotes. The sticky substance that holds bacteria together also excludes most antibiotics and disinfectants, making biofilm bacteria hardier than their planktonic counterparts. Overall, biofilms are very difficult to destroy because they are resistant to many common forms of sterilization.

Note:
Art Connection



Five stages of biofilm development are shown. During stage 1, initial attachment, bacteria adhere to a solid surface via weak van der Waals interactions. During stage 2, irreversible attachment, hairlike appendages called pili permanently anchor the bacteria to the surface. During stage 3, maturation I, the biofilm grows through cell division and recruitment of other bacteria. An extracellular matrix composed primarily of polysaccharides holds the biofilm together. During stage 4, maturation II, the biofilm continues to grow and takes on a more complex shape. During stage 5, dispersal, the

biofilm matrix is partly broken down, allowing some bacteria to escape and colonize another surface. Micrographs of a *Pseudomonas aeruginosa* biofilm in each of the stages of development are shown. (credit: D. Davis, Don Monroe, PLoS)

Compared to free-floating bacteria, bacteria in biofilms often show increased resistance to antibiotics and detergents. Why do you think this might be the case?

Section Summary

Prokaryotes existed for billions of years before plants and animals appeared. Hot springs and hydrothermal vents may have been the environments in which life began. Microbial mats are thought to represent the earliest forms of life on Earth, and there is fossil evidence of their presence about 3.5 billion years ago. A microbial mat is a multi-layered sheet of prokaryotes that grows at interfaces between different types of material, mostly on moist surfaces. During the first 2 billion years, the atmosphere was anoxic and only anaerobic organisms were able to live. Cyanobacteria evolved from early phototrophs and began the oxygenation of the atmosphere. The increase in oxygen concentration allowed the evolution of other life forms. Fossilized microbial mats are called stromatolites and consist of laminated organo-sedimentary structures formed by precipitation of minerals by prokaryotes. They represent the earliest fossil record of life on Earth.

Bacteria and archaea grow in virtually every environment. Those that survive under extreme conditions are called extremophiles (extreme lovers). Some prokaryotes cannot grow in a laboratory setting, but they are not dead. They are in the viable-but-non-culturable (VBNC) state. The VBNC state occurs when prokaryotes enter a dormant state in response to

environmental stressors. Most prokaryotes are social and prefer to live in communities where interactions take place. A biofilm is a microbial community held together in a gummy-textured matrix.

Art Connections

Exercise:

Problem:

[link] Compared to free-floating bacteria, bacteria in biofilms often show increased resistance to antibiotics and detergents. Why do you think this might be the case?

Solution:

[link] The extracellular matrix and outer layer of cells protects the inner bacteria. The close proximity of cells also facilitates lateral gene transfer, a process by which genes such as antibiotic resistance genes are transferred from one bacterium to another. And even if lateral gene transfer does not occur, one bacterium that produces an exo-enzyme that destroys antibiotic may save neighboring bacteria.

Review Questions

Exercise:

Problem:

The first forms of life on Earth were thought to be_________.

a. single-celled plants
b. prokaryotes
c. insects
d. large animals such as dinosaurs

Solution:

A

Exercise:

Problem: Microbial mats ___________.

a. are the earliest forms of life on Earth
b. obtained their energy and food from hydrothermal vents
c. are multi-layered sheet of prokaryotes including mostly bacteria but also archaea
d. all of the above

Solution:

D

Exercise:

Problem: The first organisms that oxygenated the atmosphere were

a. cyanobacteria
b. phototrophic organisms
c. anaerobic organisms
d. all of the above

Solution:

A

Exercise:

Problem: Halophiles are organisms that require_________.

a. a salt concentration of at least 0.2 M
b. high sugar concentration
c. the addition of halogens
d. all of the above

Solution:

A

Free Response

Exercise:

Problem:

Describe briefly how you would detect the presence of a non-culturable prokaryote in an environmental sample.

Solution:

As the organisms are non-culturable, the presence could be detected through molecular techniques, such as PCR.

Exercise:

Problem:

Why do scientists believe that the first organisms on Earth were extremophiles?

Solution:

Because the environmental conditions on Earth were extreme: high temperatures, lack of oxygen, high radiation, and the like.

Glossary

acidophile
: organism with optimal growth pH of three or below

alkaliphile

organism with optimal growth pH of nine or above

anaerobic
refers to organisms that grow without oxygen

anoxic
without oxygen

biofilm
a microbial community growing together on a surface, often held together with a gummy matrix

cyanobacteria
bacteria that evolved from early phototrophs and oxygenated the atmosphere; also known as blue-green algae

extremophile
organism that grows under extreme or harsh conditions

halophile
organism that require a salt concentration of at least 0.2 M

hydrothermal vent
fissure in Earth’s surface that releases geothermally heated water

hyperthermophile
organism that grows at temperatures between 80–122 °C

microbial mat
multi-layered sheet of prokaryotes that may include bacteria and archaea

nutrient
essential substances for growth, such as carbon and nitrogen

osmophile
organism that grows in a high sugar concentration

phototroph

organism that is able to make its own food by converting solar energy to chemical energy

psychrophile
: organism that grows at temperatures of -15 °C or lower

radioresistant
: organism that grows in high levels of radiation

resuscitation
: process by which prokaryotes that are in the VBNC state return to viability

stromatolite
: layered sedimentary structure formed by precipitation of minerals by prokaryotes in microbial mats

thermophile
: organism that lives at temperatures between 60–80 °C

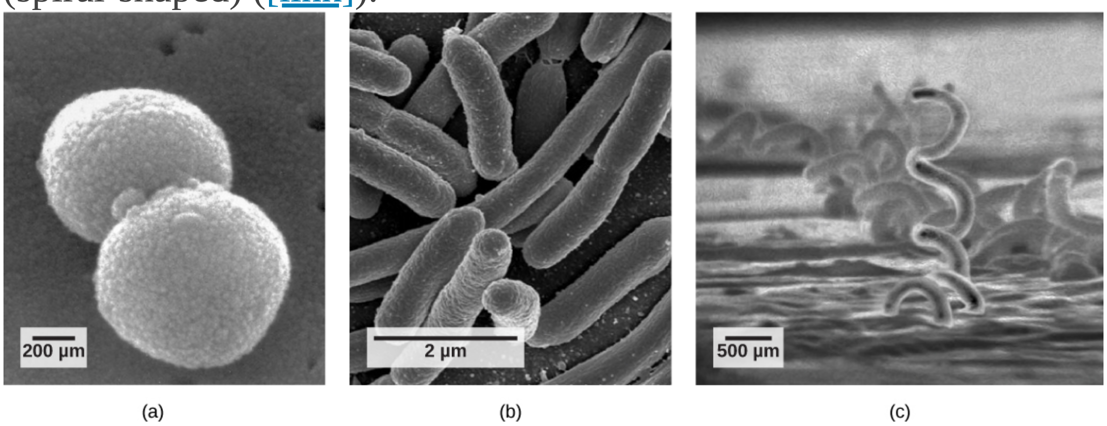
viable-but-non-culturable (VBNC) state
: survival mechanism of bacteria facing environmental stress conditions

Structure of Prokaryotes

By the end of this section, you will be able to:

- Describe the basic structure of a typical prokaryote
- Describe important differences in structure between Archaea and Bacteria

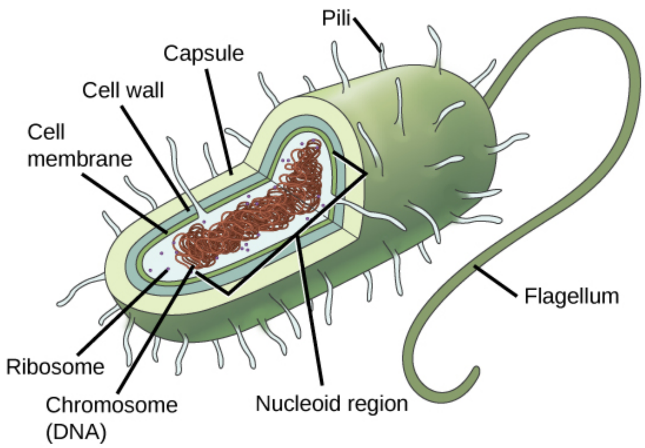
There are many differences between prokaryotic and eukaryotic cells. However, all cells have four common structures: the plasma membrane, which functions as a barrier for the cell and separates the cell from its environment; the cytoplasm, a jelly-like substance inside the cell; nucleic acids, the genetic material of the cell; and ribosomes, where protein synthesis takes place. Prokaryotes come in various shapes, but many fall into three categories: cocci (spherical), bacilli (rod-shaped), and spirilli (spiral-shaped) ([link]).



Prokaryotes fall into three basic categories based on their shape, visualized here using scanning electron microscopy: (a) cocci, or spherical (a pair is shown); (b) bacilli, or rod-shaped; and (c) spirilli, or spiral-shaped. (credit a: modification of work by Janice Haney Carr, Dr. Richard Facklam, CDC; credit c: modification of work by Dr. David Cox; scale-bar data from Matt Russell)

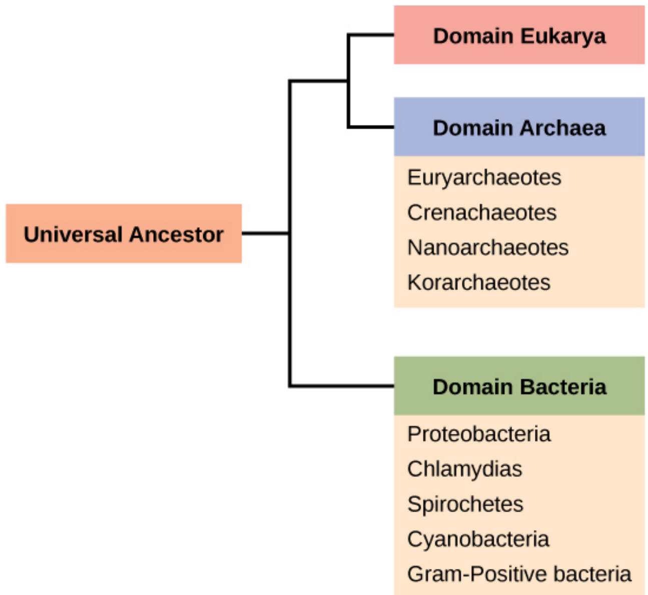
The Prokaryotic Cell

Recall that prokaryotes ([link]) are unicellular organisms that lack organelles or other internal membrane-bound structures. Therefore, they do not have a nucleus but instead generally have a single chromosome—a piece of circular, double-stranded DNA located in an area of the cell called the nucleoid. Most prokaryotes have a cell wall outside the plasma membrane.



The features of a typical prokaryotic cell are shown.

Recall that prokaryotes are divided into two different domains, Bacteria and Archaea, which together with Eukarya, comprise the three domains of life ([link]).

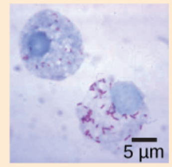
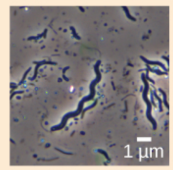
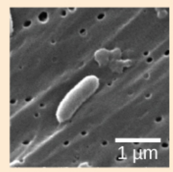
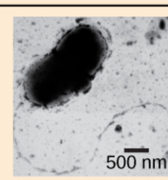
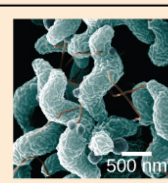


Bacteria and Archaea are both prokaryotes but differ enough to be placed in separate domains. An ancestor of modern Archaea is believed to have given rise to Eukarya, the third domain of life. Archaeal and bacterial phyla are shown; the evolutionary relationship between these phyla is still open to debate.

The composition of the cell wall differs significantly between the domains Bacteria and Archaea. The composition of their cell walls also differs from the eukaryotic cell walls found in plants (cellulose) or fungi and insects (chitin). The cell wall functions as a protective layer, and it is responsible for the organism’s shape. Some bacteria have an outer **capsule** outside the cell wall. Other structures are present in some prokaryotic species, but not in others ([link]). For example, the capsule found in some species enables the organism to attach to surfaces, protects it from dehydration and attack by phagocytic cells, and makes pathogens more resistant to our immune responses. Some species also have flagella (singular, flagellum) used for

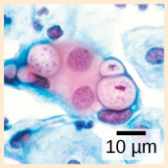
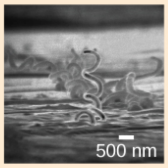
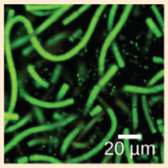
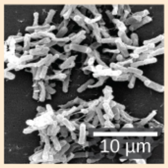
locomotion, and **pili** (singular, pilus) used for attachment to surfaces. Plasmids, which consist of extra-chromosomal DNA, are also present in many species of bacteria and archaea.

Characteristics of phyla of Bacteria are described in [link] and [link]; Archaea are described in [link].

Bacteria of Phylum Proteobacteria		
Class	**Representative organisms**	**Representative micrograph**
Alpha Proteobacteria Some species are photoautotrophic but some are symbionts of plants and animals and others are pathogens. Eukaryotic mitochondria are thought be derived from bacteria in this group.	*Rhizobium* Nitrogen-fixing endosymbiont associated with the roots of legumes *Rickettsia* Obligate intracellular parasite that causes typhus and Rocky Mountain Spotted Fever (but not rickets, which is caused by Vitamin C deficiency)	 *Rickettsia rickettsia,* stained red, grow inside a host cell.
Beta Proteobacteria This group of bacteria is diverse. Some species play an important role in the nitrogen cycle.	*Nitrosomas* Species from this group oxidize ammonia into nitrite. *Spirillum minus* Causes rat-bite fever	 *Spirillum minus*
Gamma Proteobacteria Many are beneficial symbionts that populate the human gut, but others are familiar human pathogens. Some species from this subgroup oxidize sulfur compounds.	*Escherichia coli* Normally beneficial microbe of the human gut, but some strains cause disease *Salmonella* Certain strains cause food poisoning or typhoid fever *Yersinia pestis* Causative agent of Bubonic plague *Psuedomonas aeruginosa* Causes lung infections *Vibrio cholera* Causative agent of cholera *Chromatium* Sulfur-producing bacteria that oxidize sulfur, producing H_2S	 *Vibrio cholera*
Delta Proteobacteria Some species generate a spore-forming fruiting body in adverse conditions. Others reduce sulfate and sulfur.	*Myxobacteria* Generate spore-forming fruiting bodies in adverse conditions *Desulfovibrio vulgaris* Aneorobic, sulfate-reducing bacterium	 *Desulfovibrio vulgaris*
Epsilon Proteobacteria Many species inhabit the digestive tract of animals as symbionts or pathogens. Bacteria from this group have been found in deep-sea hydrothermal vents and cold seep habitats.	*Campylobacter* Causes blood poisoning and intestinal inflammation *Heliobacter pylori* Causes stomach ulcers	 *Campylobacter*

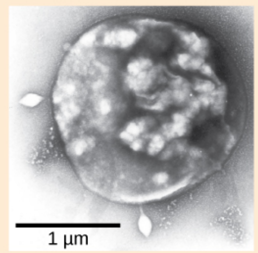
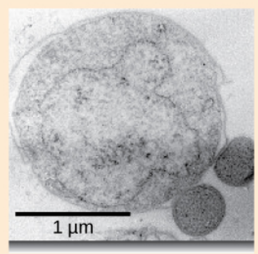
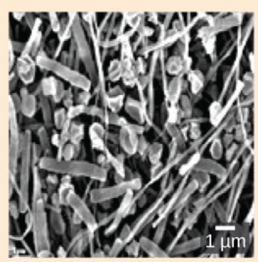
Phylum Proteobacteria is one of up to 52 bacteria phyla. Proteobacteria is further subdivided into five classes, Alpha through Epsilon. (credit “Rickettsia rickettsia”: modification of work by CDC; credit “Spirillum minus”: modification of work by Wolframm Adlassnig; credit “Vibrio cholera”: modification of work by Janice Haney Carr, CDC; credit “Desulfovibrio

vulgaris”: modification of work by Graham Bradley; credit “Campylobacter”: modification of work by De Wood, Pooley, USDA, ARS, EMU; scale-bar data from Matt Russell)

Bacteria: Chlamydia, Spirochaetae, Cyanobacteria, and Gram-positive		
Phylum	**Representative organisms**	**Representative micrograph**
Chlamydias All members of this group are obligate intracellular parasites of animal cells. Cells walls lack peptidoglycan.	*Chlamydia trachomatis* Common sexually transmitted disease that can lead to blindness	 In this pap smear, *Chlamydia trachomatis* appear as pink inclusions inside cells.
Spirochetes Most members of this species, which has spiral-shaped cells, are free-living aneaerobes, but some are pathogenic. Flagella run lengthwise in the periplasmic space between the inner and outer membrane.	*Treponema pallidum* Causative agent of syphilis *Borrelia burgdorferi* Causative agent of Lyme disease	 *Treponema pallidum*
Cyanobacteria Also known as blue-green algae, these bacteria obtain their energy through photosynthesis. They are ubiquitous, found in terrestrial, marine, and freshwater environments. Eukaryotic chloroplasts are thought be derived from bacteria in this group.	*Prochlorococcus* Believed to be the most abundant photosynthetic organism on earth; responsible for generating half the world’s oxygen	 *Phormidium*
Gram-positive Bacteria Soil-dwelling members of this subgroup decompose organic matter. Some species cause disease. They have a thick cell wall and lack an outer membrane.	*Bacillus anthracis* Causes anthrax *Clostridium botulinum* Causes Botulism *Clostridium difficile* Causes diarrhea during antibiotic therapy *Streptomyces* Many antibiotics, including streptomyocin, are derived from these bacteria. *Mycoplasmas* These tiny bacteria, the smallest known, lack a cell wall. Some are free-living, and some are pathogenic.	 *Clostridium difficile*

Chlamydia, Spirochetes, Cyanobacteria, and Gram-positive bacteria are described in this table. Note that bacterial shape is not phylum-dependent; bacteria within a phylum may be cocci,

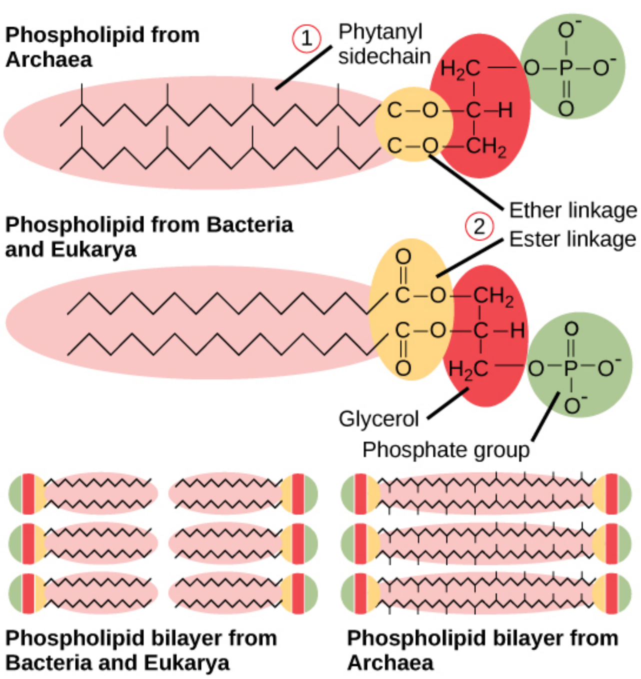
rod-shaped, or spiral. (credit “Chlamydia trachomatis”: modification of work by Dr. Lance Liotta Laboratory, NCI; credit “Treponema pallidum”: modification of work by Dr. David Cox, CDC; credit “Phormidium”: modification of work by USGS; credit “Clostridium difficile”: modification of work by Lois S. Wiggs, CDC; scale-bar data from Matt Russell)

Archaea		
Phylum	**Representative organisms**	**Representative micrograph**
Euryarchaeota This phylum includes methanogens, which produce methane as a metabolic waste product, and halobacteria, which live in an extreme saline environment.	*Methanogens* Methane production causes flatulence in humans and other animals. *Halobacteria* Large blooms of this salt-loving archaea appear reddish due to the presence of bacterirhodopsin in the membrane. Bacteriorhodopsin is related to the retinal pigment rhodopsin.	 *Halobacterium* strain NRC-1
Crenarchaeota Members of the ubiquitous phylum play an important role in the fixation of carbon. Many members of this group are sulfur-dependent extremophiles. Some are thermophilic or hyperthermophilic.	*Sulfolobus* Members of this genus grow in volcanic springs at temperatures between 75° and 80°C and at a pH between 2 and 3.	 *Sulfolobus* being infected by bacteriophage
Nanoarchaeota This group currently contains only one species, *Nanoarchaeum equitans*.	*Nanoarchaeum equitans* This species was isolated from the bottom of the Atlantic Ocean and from a hydrothermal vent at Yellowstone National Park. It is an obligate symbiont with *Ignicoccus*, another species of archaea.	 *Nanoarchaeum equitans* (small dark spheres) are in contact with their larger host, *Ignicoccus*.
Korarchaeota Members of this phylum, considered to be one of the most primitive forms of life, have only been found in the Obsidian Pool, a hot spring at Yellowstone National Park.	No members of this species have been cultivated.	 This image shows a variety of korarchaeota species from the Obsidian Pool at Yellowstone National Park.

Archaea are separated into four phyla: the Korarchaeota, Euryarchaeota, Crenarchaeota, and Nanoarchaeota. (credit “Halobacterium”: modification of work by NASA; credit “Nanoarchaeotum equitans”: modification of work by Karl O. Stetter; credit “korarchaeota”: modification of work by Office of Science of the U.S. Dept. of Energy; scale-bar data from Matt Russell)

The Plasma Membrane

The plasma membrane is a thin lipid bilayer (6 to 8 nanometers) that completely surrounds the cell and separates the inside from the outside. Its selectively permeable nature keeps ions, proteins, and other molecules within the cell and prevents them from diffusing into the extracellular environment, while other molecules may move through the membrane. Recall that the general structure of a cell membrane is a phospholipid bilayer composed of two layers of lipid molecules. In archaeal cell membranes, isoprene (phytanyl) chains linked to glycerol replace the fatty acids linked to glycerol in bacterial membranes. Some archaeal membranes are lipid monolayers instead of bilayers ([link]).



Archaeal phospholipids differ from those found in Bacteria and

Eukarya in two ways. First, they have branched phytanyl sidechains instead of linear ones. Second, an ether bond instead of an ester bond connects the lipid to the glycerol.

The Cell Wall

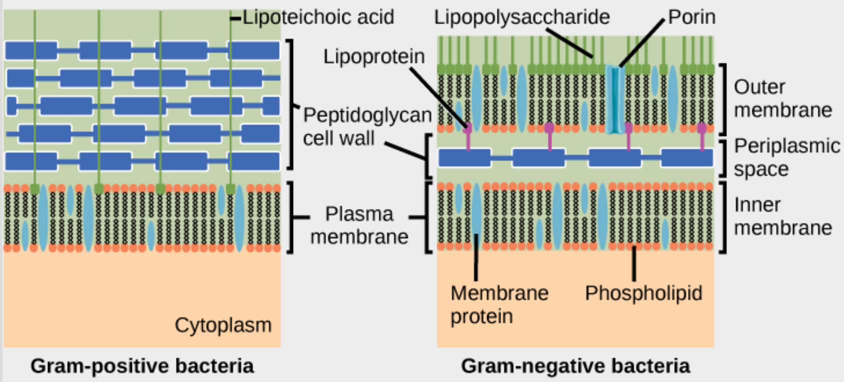
The cytoplasm of prokaryotic cells has a high concentration of dissolved solutes. Therefore, the osmotic pressure within the cell is relatively high. The cell wall is a protective layer that surrounds some cells and gives them shape and rigidity. It is located outside the cell membrane and prevents osmotic lysis (bursting due to increasing volume). The chemical composition of the cell walls varies between archaea and bacteria, and also varies between bacterial species.

Bacterial cell walls contain **peptidoglycan**, composed of polysaccharide chains that are cross-linked by unusual peptides containing both L- and D-amino acids including D-glutamic acid and D-alanine. Proteins normally have only L-amino acids; as a consequence, many of our antibiotics work by mimicking D-amino acids and therefore have specific effects on bacterial cell wall development. There are more than 100 different forms of peptidoglycan. **S-layer** (surface layer) proteins are also present on the outside of cell walls of both archaea and bacteria.

Bacteria are divided into two major groups: **Gram positive** and **Gram negative**, based on their reaction to Gram staining. Note that all Gram-positive bacteria belong to one phylum; bacteria in the other phyla (Proteobacteria, Chlamydias, Spirochetes, Cyanobacteria, and others) are Gram-negative. The Gram staining method is named after its inventor, Danish scientist Hans Christian Gram (1853–1938). The different bacterial responses to the staining procedure are ultimately due to cell wall structure. Gram-positive organisms typically lack the outer membrane found in Gram-negative organisms ([link]). Up to 90 percent of the cell wall in Gram-

positive bacteria is composed of peptidoglycan, and most of the rest is composed of acidic substances called **teichoic acids**. Teichoic acids may be covalently linked to lipids in the plasma membrane to form lipoteichoic acids. Lipoteichoic acids anchor the cell wall to the cell membrane. Gram-negative bacteria have a relatively thin cell wall composed of a few layers of peptidoglycan (only 10 percent of the total cell wall), surrounded by an outer envelope containing lipopolysaccharides (LPS) and lipoproteins. This outer envelope is sometimes referred to as a second lipid bilayer. The chemistry of this outer envelope is very different, however, from that of the typical lipid bilayer that forms plasma membranes.

Note:
Art Connection



Bacteria are divided into two major groups: Gram positive and Gram negative. Both groups have a cell wall composed of peptidoglycan: in Gram-positive bacteria, the wall is thick, whereas in Gram-negative bacteria, the wall is thin. In Gram-negative bacteria, the cell wall is surrounded by an outer membrane that contains lipopolysaccharides and lipoproteins. Porins are proteins in this cell membrane that allow substances to pass through the outer membrane of Gram-negative bacteria. In Gram-positive bacteria, lipoteichoic acid anchors the cell wall to the cell membrane.

(credit: modification of work by "Franciscosp2"/Wikimedia Commons)

Which of the following statements is true?

a. Gram-positive bacteria have a single cell wall anchored to the cell membrane by lipoteichoic acid.
b. Porins allow entry of substances into both Gram-positive and Gram-negative bacteria.
c. The cell wall of Gram-negative bacteria is thick, and the cell wall of Gram-positive bacteria is thin.
d. Gram-negative bacteria have a cell wall made of peptidoglycan, whereas Gram-positive bacteria have a cell wall made of lipoteichoic acid.

Archaean cell walls do not have peptidoglycan. There are four different types of Archaean cell walls. One type is composed of **pseudopeptidoglycan**, which is similar to peptidoglycan in morphology but contains different sugars in the polysaccharide chain. The other three types of cell walls are composed of polysaccharides, glycoproteins, or pure protein.

Structural Differences and Similarities between Bacteria and Archaea		
Structural Characteristic	**Bacteria**	**Archaea**

Structural Differences and Similarities between Bacteria and Archaea		
Structural Characteristic	**Bacteria**	**Archaea**
Cell type	Prokaryotic	Prokaryotic
Cell morphology	Variable	Variable
Cell wall	Contains peptidoglycan	Does not contain peptidoglycan
Cell membrane type	Lipid bilayer	Lipid bilayer or lipid monolayer
Plasma membrane lipids	Fatty acids	Phytanyl groups

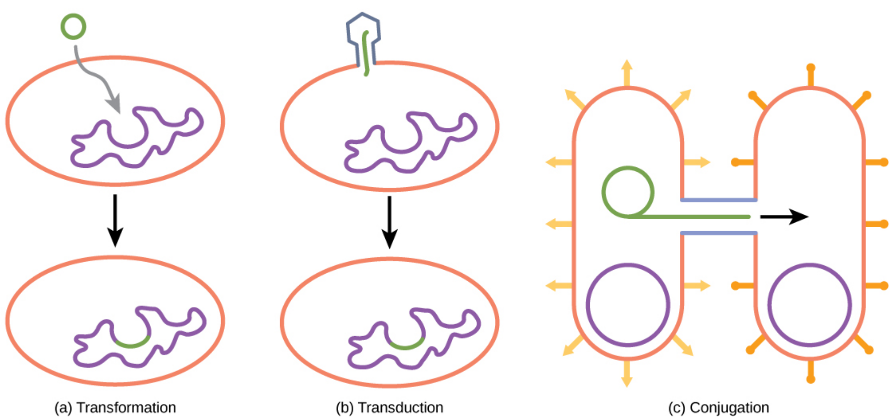
Reproduction

Reproduction in prokaryotes is asexual and usually takes place by binary fission. Recall that the DNA of a prokaryote exists as a single, circular chromosome. Prokaryotes do not undergo mitosis. Rather the chromosome is replicated and the two resulting copies separate from one another, due to the growth of the cell. The prokaryote, now enlarged, is pinched inward at its equator and the two resulting cells, which are clones, separate. Binary fission does not provide an opportunity for genetic recombination or genetic diversity, but prokaryotes can share genes by three other mechanisms.

In **transformation**, the prokaryote takes in DNA found in its environment that is shed by other prokaryotes. If a nonpathogenic bacterium takes up DNA for a toxin gene from a pathogen and incorporates the new DNA into its own chromosome, it too may become pathogenic. In **transduction**, bacteriophages, the viruses that infect bacteria, sometimes also move short pieces of chromosomal DNA from one bacterium to another. Transduction

results in a recombinant organism. Archaea are not affected by bacteriophages but instead have their own viruses that translocate genetic material from one individual to another. In **conjugation**, DNA is transferred from one prokaryote to another by means of a pilus, which brings the organisms into contact with one another. The DNA transferred can be in the form of a plasmid or as a hybrid, containing both plasmid and chromosomal DNA. These three processes of DNA exchange are shown in [link].

Reproduction can be very rapid: a few minutes for some species. This short generation time coupled with mechanisms of genetic recombination and high rates of mutation result in the rapid evolution of prokaryotes, allowing them to respond to environmental changes (such as the introduction of an antibiotic) very quickly.



Besides binary fission, there are three other mechanisms by which prokaryotes can exchange DNA. In (a) transformation, the cell takes up prokaryotic DNA directly from the environment. The DNA may remain separate as plasmid DNA or be incorporated into the host genome. In (b) transduction, a bacteriophage injects DNA into the cell that contains a small fragment of DNA from a different prokaryote. In (c) conjugation, DNA is transferred from one cell to another via a mating bridge that connects the two cells after the

sex pilus draws the two bacteria close enough to form the bridge.

Note:

Evolution Connection

The Evolution of Prokaryotes

How do scientists answer questions about the evolution of prokaryotes? Unlike with animals, artifacts in the fossil record of prokaryotes offer very little information. Fossils of ancient prokaryotes look like tiny bubbles in rock. Some scientists turn to genetics and to the principle of the molecular clock, which holds that the more recently two species have diverged, the more similar their genes (and thus proteins) will be. Conversely, species that diverged long ago will have more genes that are dissimilar.

Scientists at the NASA Astrobiology Institute and at the European Molecular Biology Laboratory collaborated to analyze the molecular evolution of 32 specific proteins common to 72 species of prokaryotes. [footnote] The model they derived from their data indicates that three important groups of bacteria—Actinobacteria, *Deinococcus*, and Cyanobacteria (which the authors call *Terrabacteria*)—were the first to colonize land. (Recall that *Deinococcus* is a genus of prokaryote—a bacterium—that is highly resistant to ionizing radiation.) Cyanobacteria are photosynthesizers, while Actinobacteria are a group of very common bacteria that include species important in decomposition of organic wastes.

Battistuzzi, FU, Feijao, A, and Hedges, SB. A genomic timescale of prokaryote evolution: Insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BioMed Central: Evolutionary Biology* 4 (2004): 44, doi:10.1186/1471-2148-4-44.

The timelines of divergence suggest that bacteria (members of the domain Bacteria) diverged from common ancestral species between 2.5 and 3.2 billion years ago, whereas archaea diverged earlier: between 3.1 and 4.1 billion years ago. Eukarya later diverged off the Archaean line. The work further suggests that stromatolites that formed prior to the advent of cyanobacteria (about 2.6 billion years ago) photosynthesized in an anoxic environment and that because of the modifications of the Terrabacteria for

land (resistance to drying and the possession of compounds that protect the organism from excess light), photosynthesis using oxygen may be closely linked to adaptations to survive on land.

Section Summary

Prokaryotes (domains Archaea and Bacteria) are single-celled organisms lacking a nucleus. They have a single piece of circular DNA in the nucleoid area of the cell. Most prokaryotes have a cell wall that lies outside the boundary of the plasma membrane. Some prokaryotes may have additional structures such as a capsule, flagella, and pili. Bacteria and Archaea differ in the lipid composition of their cell membranes and the characteristics of the cell wall. In archaeal membranes, phytanyl units, rather than fatty acids, are linked to glycerol. Some archaeal membranes are lipid monolayers instead of bilayers.

The cell wall is located outside the cell membrane and prevents osmotic lysis. The chemical composition of cell walls varies between species. Bacterial cell walls contain peptidoglycan. Archaean cell walls do not have peptidoglycan, but they may have pseudopeptidoglycan, polysaccharides, glycoproteins, or protein-based cell walls. Bacteria can be divided into two major groups: Gram positive and Gram negative, based on the Gram stain reaction. Gram-positive organisms have a thick cell wall, together with teichoic acids. Gram-negative organisms have a thin cell wall and an outer envelope containing lipopolysaccharides and lipoproteins.

Art Connections

Exercise:

Problem: [link] Which of the following statements is true?

a. Gram-positive bacteria have a single cell wall anchored to the cell membrane by lipoteichoic acid.

b. Porins allow entry of substances into both Gram-positive and Gram-negative bacteria.
c. The cell wall of Gram-negative bacteria is thick, and the cell wall of Gram-positive bacteria is thin.
d. Gram-negative bacteria have a cell wall made of peptidoglycan, whereas Gram-positive bacteria have a cell wall made of lipoteichoic acid.

Solution:

[link] A

Review Questions

Exercise:

Problem:

The presence of a membrane-enclosed nucleus is a characteristic of _________.

a. prokaryotic cells
b. eukaryotic cells
c. all cells
d. viruses

Solution:

B

Exercise:

Problem: Which of the following consist of prokaryotic cells?

a. bacteria and fungi
b. archaea and fungi

c. protists and animals
d. bacteria and archaea

Solution:

D

Exercise:

Problem: The cell wall is ________.

a. interior to the cell membrane
b. exterior to the cell membrane
c. a part of the cell membrane
d. interior or exterior, depending on the particular cell

Solution:

B

Exercise:

Problem:

Organisms most likely to be found in extreme environments are ________.

a. fungi
b. bacteria
c. viruses
d. archaea

Solution:

B

Exercise:

Problem:

Prokaryotes stain as Gram-positive or Gram-negative because of differences in the cell _______.

a. wall
b. cytoplasm
c. nucleus
d. chromosome

Solution:

A

Exercise:

Problem:

Pseudopeptidoglycan is a characteristic of the walls of ________.

a. eukaryotic cells
b. bacterial prokaryotic cells
c. archaean prokaryotic cells
d. bacterial and archaean prokaryotic cells

Solution:

C

Exercise:

Problem:

The lipopolysaccharide layer (LPS) is a characteristic of the wall of ________.

a. archaean cells
b. Gram-negative bacteria

c. bacterial prokaryotic cells
d. eukaryotic cells

Solution:

B

Free Response

Exercise:

Problem: Mention three differences between bacteria and archaea.

Solution:

Responses will vary. A possible answer is: Bacteria contain peptidoglycan in the cell wall; archaea do not. The cell membrane in bacteria is a lipid bilayer; in archaea, it can be a lipid bilayer or a monolayer. Bacteria contain fatty acids on the cell membrane, whereas archaea contain phytanyl.

Exercise:

Problem:

Explain the statement that both types, bacteria and archaea, have the same basic structures, but built from different chemical components.

Solution:

Both bacteria and archaea have cell membranes and they both contain a hydrophobic portion. In the case of bacteria, it is a fatty acid; in the case of archaea, it is a hydrocarbon (phytanyl). Both bacteria and archaea have a cell wall that protects them. In the case of bacteria, it is composed of peptidoglycan, whereas in the case of archaea, it is pseudopeptidoglycan, polysaccharides, glycoproteins, or pure protein. Bacterial and archaeal flagella also differ in their chemical structure.

Glossary

capsule
: external structure that enables a prokaryote to attach to surfaces and protects it from dehydration

conjugation
: process by which prokaryotes move DNA from one individual to another using a pilus

Gram negative
: bacterium whose cell wall contains little peptidoglycan but has an outer membrane

Gram positive
: bacterium that contains mainly peptidoglycan in its cell walls

peptidoglycan
: material composed of polysaccharide chains cross-linked to unusual peptides

pilus
: surface appendage of some prokaryotes used for attachment to surfaces including other prokaryotes

pseudopeptidoglycan
: component of archaea cell walls that is similar to peptidoglycan in morphology but contains different sugars

S-layer
: surface-layer protein present on the outside of cell walls of archaea and bacteria

teichoic acid
: polymer associated with the cell wall of Gram-positive bacteria

transduction

process by which a bacteriophage moves DNA from one prokaryote to another

transformation
: process by which a prokaryote takes in DNA found in its environment that is shed by other prokaryotes

Prokaryotic Metabolism
By the end of this section, you will be able to:

- Identify the macronutrients needed by prokaryotes, and explain their importance
- Describe the ways in which prokaryotes get energy and carbon for life processes
- Describe the roles of prokaryotes in the carbon and nitrogen cycles

Prokaryotes are metabolically diverse organisms. There are many different environments on Earth with various energy and carbon sources, and variable conditions. Prokaryotes have been able to live in every environment by using whatever energy and carbon sources are available. Prokaryotes fill many niches on Earth, including being involved in nutrient cycles such as nitrogen and carbon cycles, decomposing dead organisms, and thriving inside living organisms, including humans. The very broad range of environments that prokaryotes occupy is possible because they have diverse metabolic processes.

Needs of Prokaryotes

The diverse environments and ecosystems on Earth have a wide range of conditions in terms of temperature, available nutrients, acidity, salinity, and energy sources. Prokaryotes are very well equipped to make their living out of a vast array of nutrients and conditions. To live, prokaryotes need a source of energy, a source of carbon, and some additional nutrients.

Macronutrients

Cells are essentially a well-organized assemblage of macromolecules and water. Recall that macromolecules are produced by the polymerization of smaller units called monomers. For cells to build all of the molecules required to sustain life, they need certain substances, collectively called **nutrients**. When prokaryotes grow in nature, they obtain their nutrients from the environment. Nutrients that are required in large amounts are called macronutrients, whereas those required in smaller or trace amounts are called micronutrients. Just a handful of elements are considered macronutrients—carbon, hydrogen, oxygen, nitrogen, phosphorus, and sulfur. (A mnemonic for remembering these elements is the acronym *CHONPS*.)

Why are these macronutrients needed in large amounts? They are the components of organic compounds in cells, including water. Carbon is the major element in all macromolecules: carbohydrates, proteins, nucleic acids, lipids, and many other compounds. Carbon accounts for about 50 percent of the composition of the cell. Nitrogen represents 12 percent of the total dry weight of a typical cell and is a component of proteins, nucleic acids, and other cell constituents. Most of the nitrogen available in nature is either atmospheric nitrogen (N_2) or another inorganic form. Diatomic (N_2) nitrogen, however, can be converted into an organic form only by certain organisms, called nitrogen-fixing organisms. Both hydrogen and oxygen are part of many organic compounds and of water. Phosphorus is required by all organisms for the synthesis of nucleotides and phospholipids. Sulfur is part of the structure of some amino acids such as cysteine and methionine, and is also present in several vitamins and

coenzymes. Other important macronutrients are potassium (K), magnesium (Mg), calcium (Ca), and sodium (Na). Although these elements are required in smaller amounts, they are very important for the structure and function of the prokaryotic cell.

Micronutrients

In addition to these macronutrients, prokaryotes require various metallic elements in small amounts. These are referred to as micronutrients or trace elements. For example, iron is necessary for the function of the cytochromes involved in electron-transport reactions. Some prokaryotes require other elements—such as boron (B), chromium (Cr), and manganese (Mn)—primarily as enzyme cofactors.

The Ways in Which Prokaryotes Obtain Energy

Prokaryotes can use different sources of energy to assemble macromolecules from smaller molecules. **Phototrophs** (or phototrophic organisms) obtain their energy from sunlight. **Chemotrophs** (or chemosynthetic organisms) obtain their energy from chemical compounds. Chemotrophs that can use organic compounds as energy sources are called chemoorganotrophs. Those that can also use inorganic compounds as energy sources are called chemolitotrophs.

The Ways in Which Prokaryotes Obtain Carbon

Prokaryotes not only can use different sources of energy but also different sources of carbon compounds. Recall that organisms that are able to fix inorganic carbon are called autotrophs. Autotrophic prokaryotes synthesize organic molecules from carbon dioxide. In contrast, heterotrophic prokaryotes obtain carbon from organic compounds. To make the picture more complex, the terms that describe how prokaryotes obtain energy and carbon can be combined. Thus, photoautotrophs use energy from sunlight, and carbon from carbon dioxide and water, whereas chemoheterotrophs obtain energy and carbon from an organic chemical source. Chemolitoautotrophs obtain their energy from inorganic compounds, and they build their complex molecules from carbon dioxide. The table below ([link]) summarizes carbon and energy sources in prokaryotes.

Carbon and Energy Sources in Prokaryotes

Carbon and Energy Sources in Prokaryotes

Energy Sources			Carbon Sources	
Light	Chemicals		Carbon dioxide	Organic compounds
Phototrophs	Chemotrophs		Autotrophs	Heterotrophs
	Organic chemicals	Inorganic chemicals		
	Chemo-organotrophs	Chemolithotrophs		

Role of Prokaryotes in Ecosystems

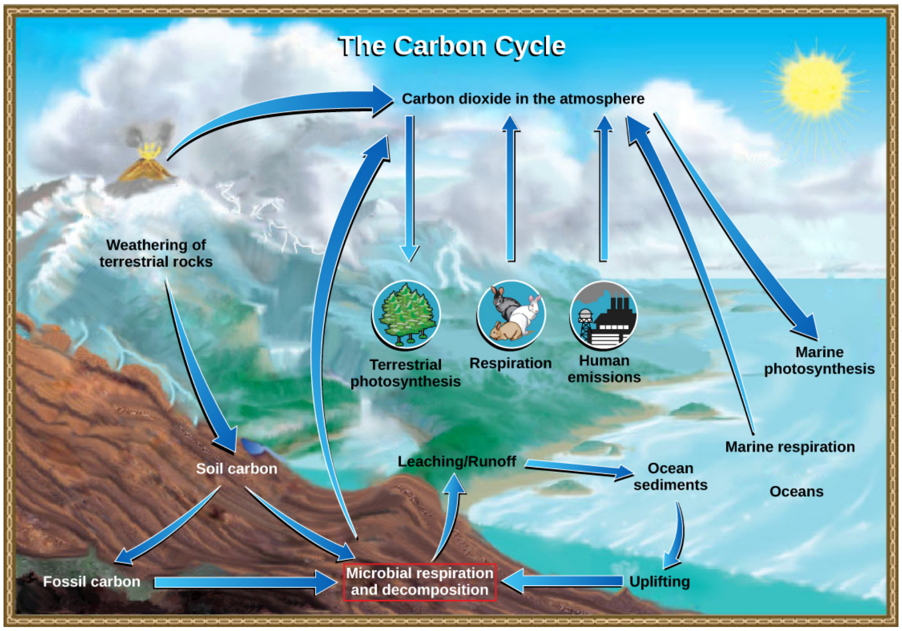
Prokaryotes are ubiquitous: There is no niche or ecosystem in which they are not present. Prokaryotes play many roles in the environments they occupy. The roles they play in the carbon and nitrogen cycles are vital to life on Earth.

Prokaryotes and the Carbon Cycle

Carbon is one of the most important macronutrients, and prokaryotes play an important role in the carbon cycle ([link]). Carbon is cycled through Earth's major reservoirs: land, the atmosphere, aquatic environments, sediments and rocks, and biomass. The movement of carbon is via carbon dioxide, which is removed from the atmosphere by land plants and marine prokaryotes, and is returned to the atmosphere via the respiration of chemoorganotrophic organisms, including prokaryotes, fungi, and animals. Although the largest carbon reservoir in terrestrial ecosystems is in rocks and sediments, that carbon is not readily available.

A large amount of available carbon is found in land plants. Plants, which are producers, use carbon dioxide from the air to synthesize carbon compounds. Related to this, one very significant source of carbon compounds is humus, which is a mixture of organic materials from dead plants and prokaryotes that have resisted decomposition. Consumers such as animals use organic compounds generated by producers and release carbon dioxide to the atmosphere. Then, bacteria and fungi, collectively called **decomposers**, carry out the breakdown (decomposition) of plants and animals and their organic compounds. The most important contributor of carbon dioxide to the atmosphere is microbial decomposition of dead material (dead animals, plants, and humus) that undergo respiration.

In aqueous environments and their anoxic sediments, there is another carbon cycle taking place. In this case, the cycle is based on one-carbon compounds. In anoxic sediments, prokaryotes, mostly archaea, produce methane (CH_4). This methane moves into the zone above the sediment, which is richer in oxygen and supports bacteria called methane oxidizers that oxidize methane to carbon dioxide, which then returns to the atmosphere.



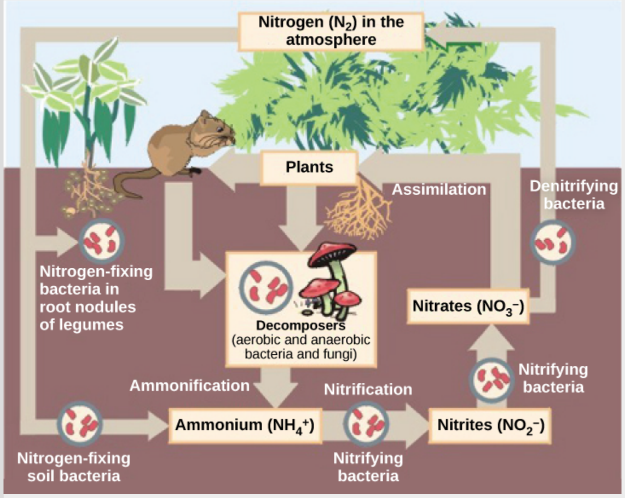
Prokaryotes play a significant role in continuously moving carbon through the biosphere. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Prokaryotes and the Nitrogen Cycle

Nitrogen is a very important element for life because it is part of proteins and nucleic acids. It is a macronutrient, and in nature, it is recycled from organic compounds to ammonia, ammonium ions, nitrate, nitrite, and nitrogen gas by myriad processes, many of which are carried out only by prokaryotes. As illustrated in [link], prokaryotes are key to the nitrogen cycle. The largest pool of nitrogen available in the terrestrial ecosystem is gaseous nitrogen from the air, but this nitrogen is not usable by plants, which are primary producers. Gaseous nitrogen is transformed, or “fixed” into more readily available forms such as ammonia through the process of **nitrogen fixation**. Ammonia can be used by plants or converted to other forms.

Another source of ammonia is **ammonification**, the process by which ammonia is released during the decomposition of nitrogen-containing organic compounds. Ammonia released to the atmosphere, however, represents only 15 percent of the total nitrogen released; the rest is as N_2 and N_2O. Ammonia is catabolized anaerobically by some prokaryotes, yielding N_2 as the final product. **Nitrification** is the conversion of ammonium to nitrite and nitrate. Nitrification in soils is carried out by bacteria belonging to the genera *Nitrosomas*, *Nitrobacter*, and *Nitrospira*. The bacteria performs the reverse process, the reduction of nitrate from the soils to gaseous compounds such as N_2O, NO, and N_2, a process called **denitrification**.

Note:
Art Connection



Prokaryotes play a key role in the nitrogen cycle. (credit: Environmental Protection Agency)

Which of the following statements about the nitrogen cycle is false?

a. Nitrogen fixing bacteria exist on the root nodules of legumes and in the soil.
b. Denitrifying bacteria convert nitrates (NO_3^-) into nitrogen gas (N_2).
c. Ammonification is the process by which ammonium ion (NH_4^+) is released from decomposing organic compounds.
d. Nitrification is the process by which nitrites (NO_2^-) are converted to ammonium ion (NH_4^+).

Section Summary

Prokaryotes are the most metabolically diverse organisms; they flourish in many different environments with various carbon energy and carbon sources, variable temperature, pH, pressure, and water availability. Nutrients required in large amounts are called macronutrients, whereas those required in trace amounts are called micronutrients or trace elements. Macronutrients include C, H, O, N, P, S, K, Mg, Ca, and Na. In addition to these macronutrients, prokaryotes require various metallic elements for growth and enzyme function. Prokaryotes use different sources of energy to assemble macromolecules from smaller molecules. Phototrophs obtain their energy from sunlight, whereas chemotrophs obtain energy from chemical compounds.

Prokaryotes play roles in the carbon and nitrogen cycles. Carbon is returned to the atmosphere by the respiration of animals and other chemoorganotrophic organisms. Consumers use organic compounds generated by producers and release carbon dioxide into the atmosphere. The most important contributor of carbon dioxide to the atmosphere is microbial decomposition of dead material. Nitrogen is recycled in nature from organic compounds to ammonia, ammonium ions, nitrite, nitrate, and nitrogen gas. Gaseous nitrogen is transformed into ammonia through nitrogen fixation. Ammonia is anaerobically catabolized by some prokaryotes, yielding N_2 as the final product. Nitrification is the conversion of ammonium into nitrite. Nitrification in soils is carried out by bacteria. Denitrification is also performed by bacteria and transforms nitrate from soils into gaseous nitrogen compounds, such as N_2O, NO, and N_2.

Art Connections

Exercise:

Problem: [link] Which of the following statements about the nitrogen cycle is false?

a. Nitrogen fixing bacteria exist on the root nodules of legumes and in the soil.
b. Denitrifying bacteria convert nitrates (NO_3^-) into nitrogen gas (N_2).
c. Ammonification is the process by which ammonium ion (NH_4^+) is released from decomposing organic compounds.
d. Nitrification is the process by which nitrites (NO_2^-) are converted to ammonium ion (NH_4^+).

Solution:

[link] D

Review Questions

Exercise:

Problem: Which of the following elements is *not* a micronutrient?

a. boron
b. calcium
c. chromium
d. manganese

Solution:

B

Exercise:

Problem:

Prokaryotes that obtain their energy from chemical compounds are called ______.

a. phototrophs
b. auxotrophs
c. chemotrophs
d. lithotrophs

Solution:

C

Exercise:

Problem: Ammonification is the process by which ______.

a. ammonia is released during the decomposition of nitrogen-containing organic compounds
b. ammonium is converted to nitrite and nitrate in soils
c. nitrate from soil is transformed to gaseous nitrogen compounds such as NO, N_2O, and N_2
d. gaseous nitrogen is fixed to yield ammonia

Solution:

A

Exercise:

Problem: Plants use carbon dioxide from the air and are therefore called ______.

a. consumers
b. producers
c. decomposer

d. carbon fixers

Solution:

B

Free Response

Exercise:

Problem:

Think about the conditions (temperature, light, pressure, and organic and inorganic materials) that you may find in a deep-sea hydrothermal vent. What type of prokaryotes, in terms of their metabolic needs (autotrophs, phototrophs, chemotrophs, etc.), would you expect to find there?

Solution:

Responses will vary. In a deep-sea hydrothermal vent, there is no light, so prokaryotes would be chemotrophs instead of phototrophs. The source of carbon would be carbon dioxide dissolved in the ocean, so they would be autotrophs. There is not a lot of organic material in the ocean, so prokaryotes would probably use inorganic sources, thus they would be chemolitotrophs. The temperatures are very high in the hydrothermal vent, so the prokaryotes would be thermophilic.

Glossary

ammonification
: process by which ammonia is released during the decomposition of nitrogen-containing organic compounds

chemotroph
: organism that obtains energy from chemical compounds

decomposer
: organism that carries out the decomposition of dead organisms

denitrification
: transformation of nitrate from soil to gaseous nitrogen compounds such as N_2O, NO and N_2

nitrification
: conversion of ammonium into nitrite and nitrate in soils

nitrogen fixation

process by which gaseous nitrogen is transformed, or “fixed” into more readily available forms such as ammonia

Bacterial Diseases in Humans

By the end of this section, you will be able to:

- Identify bacterial diseases that caused historically important plagues and epidemics
- Describe the link between biofilms and foodborne diseases
- Explain how overuse of antibiotic may be creating “super bugs”
- Explain the importance of MRSA with respect to the problems of antibiotic resistance

Devastating pathogen-borne diseases and plagues, both viral and bacterial in nature, have affected humans since the beginning of human history. The true cause of these diseases was not understood at the time, and some people thought that diseases were a spiritual punishment. Over time, people came to realize that staying apart from afflicted persons, and disposing of the corpses and personal belongings of victims of illness, reduced their own chances of getting sick.

Epidemiologists study how diseases affect a population. An **epidemic** is a disease that occurs in an unusually high number of individuals in a population at the same time. A **pandemic** is a widespread, usually worldwide, epidemic. An **endemic disease** is a disease that is constantly present, usually at low incidence, in a population.

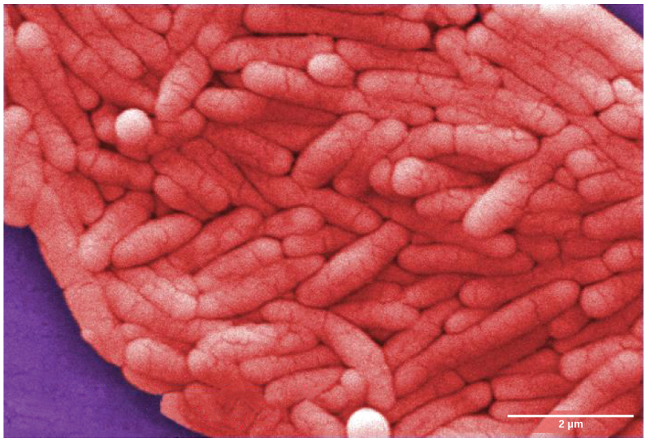
Long History of Bacterial Disease

There are records about infectious diseases as far back as 3000 B.C. A number of significant pandemics caused by bacteria have been documented over several hundred years. Some of the most memorable pandemics led to the decline of cities and nations.

In the 21st century, infectious diseases remain among the leading causes of death worldwide, despite advances made in medical research and treatments in recent decades. A disease spreads when the pathogen that causes it is passed from one person to another. For a pathogen to cause disease, it must be able to reproduce in the host’s body and damage the host in some way.

The Plague of Athens

In 430 B.C., the Plague of Athens killed one-quarter of the Athenian troops that were fighting in the great Peloponnesian War and weakened Athens' dominance and power. The plague impacted people living in overcrowded Athens as well as troops aboard ships that had to return to Athens. The source of the plague may have been identified recently when researchers from the University of Athens were able to use DNA from teeth recovered from a mass grave. The scientists identified nucleotide sequences from a pathogenic bacterium, *Salmonella enterica* serovar Typhi ([link]), which causes typhoid fever.[footnote] This disease is commonly seen in overcrowded areas and has caused epidemics throughout recorded history. Papagrigorakis MJ, Synodinos PN, and Yapijakis C. Ancient typhoid epidemic reveals possible ancestral strain of *Salmonella enterica* serovar Typhi. *Infect Genet Evol* 7 (2007): 126–7, Epub 2006 Jun.



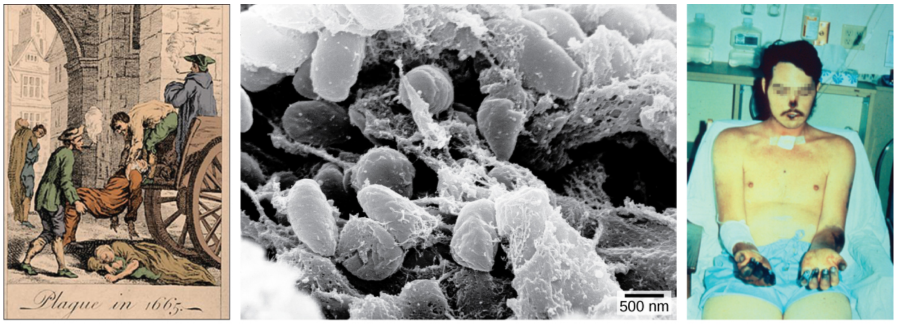
Salmonella enterica serovar Typhi, the causative agent of Typhoid fever, is a Gram-negative, rod-shaped gamma protobacterium. Typhoid fever, which is spread through feces, causes intestinal hemorrhage, high fever, delirium and dehydration. Today, between 16 and 33 million

cases of this re-emerging disease occur annually, resulting in over 200,000 deaths. Carriers of the disease can be asymptomatic. In a famous case in the early 1900s, a cook named Mary Mallon unknowingly spread the disease to over fifty people, three of whom died. Other *Salmonella* serotypes cause food poisoning. (credit: modification of work by NCI, CDC)

Bubonic Plagues

From 541 to 750, an outbreak of what was likely a bubonic plague (the Plague of Justinian), eliminated one-quarter to one-half of the human population in the eastern Mediterranean region. The population in Europe dropped by 50 percent during this outbreak. Bubonic plague would strike Europe more than once.

One of the most devastating pandemics was the **Black Death** (1346 to 1361) that is believed to have been another outbreak of bubonic plague caused by the bacterium *Yersinia pestis*. It is thought to have originated initially in China and spread along the Silk Road, a network of land and sea trade routes, to the Mediterranean region and Europe, carried by rat fleas living on black rats that were always present on ships. The Black Death reduced the world's population from an estimated 450 million to about 350 to 375 million. Bubonic plague struck London hard again in the mid-1600s ([link]). In modern times, approximately 1,000 to 3,000 cases of plague arise globally each year. Although contracting bubonic plague before antibiotics meant almost certain death, the bacterium responds to several types of modern antibiotics, and mortality rates from plague are now very low.



(a) (b) (c)

The (a) Great Plague of London killed an estimated 200,000 people, or about twenty percent of the city's population. The causative agent, the (b) bacterium *Yersinia pestis*, is a Gram-negative, rod-shaped bacterium from the class Gamma Proteobacteria. The disease is transmitted through the bite of an infected flea, which is infected by a rodent. Symptoms include swollen lymph nodes, fever, seizure, vomiting of blood, and (c) gangrene. (credit b: Rocky Mountain Laboratories, NIAID, NIH; scale-bar data from Matt Russell; credit c: Textbook of Military Medicine, Washington, D.C., U.S. Dept. of the Army, Office of the Surgeon General, Borden Institute)

Note:
Link to Learning

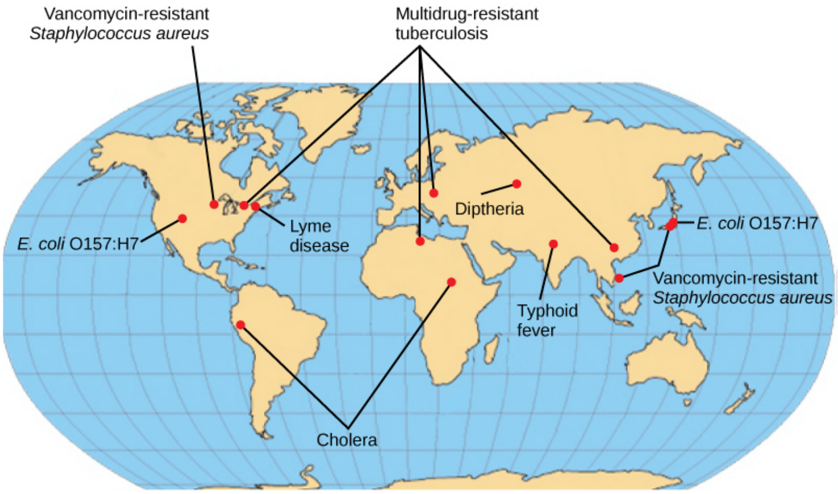
Watch a video on the modern understanding of the Black Death—bubonic plague in Europe during the 14^{th} century.

Migration of Diseases to New Populations

Over the centuries, Europeans tended to develop genetic immunity to endemic infectious diseases, but when European conquerors reached the western hemisphere, they brought with them disease-causing bacteria and viruses, which triggered epidemics that completely devastated populations of Native Americans, who had no natural resistance to many European diseases. It has been estimated that up to 90 percent of Native Americans died from infectious diseases after the arrival of Europeans, making conquest of the New World a foregone conclusion.

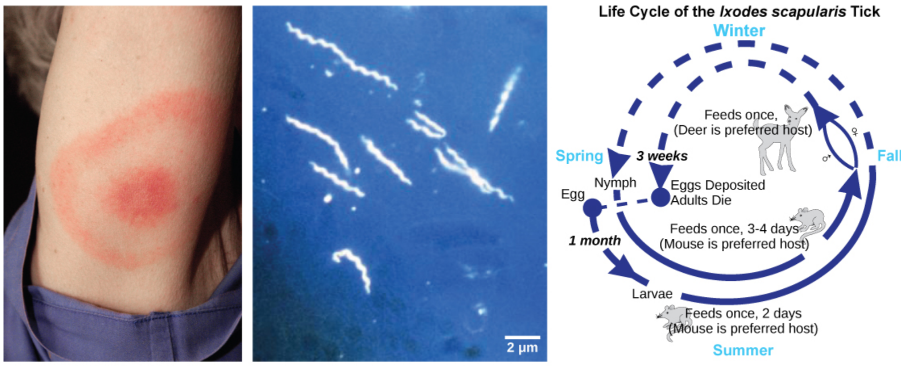
Emerging and Re-emerging Diseases

The distribution of a particular disease is dynamic. Therefore, changes in the environment, the pathogen, or the host population can dramatically impact the spread of a disease. According to the World Health Organization (WHO) an **emerging disease** ([link]) is one that has appeared in a population for the first time, or that may have existed previously but is rapidly increasing in incidence or geographic range. This definition also includes re-emerging diseases that were previously under control. Approximately 75 percent of recently emerging infectious diseases affecting humans are zoonotic diseases, **zoonoses**, diseases that primarily infect animals and are transmitted to humans; some are of viral origin and some are of bacterial origin. Brucellosis is an example of a prokaryotic zoonosis that is re-emerging in some regions, and necrotizing fasciitis (commonly known as flesh-eating bacteria) has been increasing in virulence for the last 80 years for unknown reasons.



The map shows regions where bacterial diseases are emerging or reemerging. (credit: modification of work by NIH)

Some of the present emerging diseases are not actually new, but are diseases that were catastrophic in the past ([link]). They devastated populations and became dormant for a while, just to come back, sometimes more virulent than before, as was the case with bubonic plague. Other diseases, like tuberculosis, were never eradicated but were under control in some regions of the world until coming back, mostly in urban centers with high concentrations of immunocompromised people. The WHO has identified certain diseases whose worldwide re-emergence should be monitored. Among these are two viral diseases (dengue fever and yellow fever), and three bacterial diseases (diphtheria, cholera, and bubonic plague). The war against infectious diseases has no foreseeable end.



(a) (b) (c)

Lyme disease often, but not always, results in (a) a characteristic bullseye rash. The disease is caused by a (b) Gram-negative spirochete bacterium of the genus *Borrelia*. The bacteria (c) infect ticks, which in turns infect mice. Deer are the preferred secondary host, but the ticks also may feed on humans. Untreated, the disease causes chronic disorders in the nervous system, eyes, joints, and heart. The disease is named after Lyme, Connecticut, where an outbreak occurred in 1995 and has subsequently spread. The disease is not new, however. Genetic evidence suggests that Ötzi the Iceman, a 5,300-year-old mummy found in the Alps, was infected with *Borrelia*. (credit a: James Gathany, CDC; credit b: CDC; scale-bar data from Matt Russell)

Biofilms and Disease

Recall that biofilms are microbial communities that are very difficult to destroy. They are responsible for diseases such as infections in patients with cystic fibrosis, Legionnaires’ disease, and otitis media. They produce dental plaque and colonize catheters, prostheses, transcutaneous and orthopedic devices, contact lenses, and internal devices such as pacemakers. They also form in open wounds and burned tissue. In healthcare environments,

biofilms grow on hemodialysis machines, mechanical ventilators, shunts, and other medical equipment. In fact, 65 percent of all infections acquired in the hospital (nosocomial infections) are attributed to biofilms. Biofilms are also related to diseases contracted from food because they colonize the surfaces of vegetable leaves and meat, as well as food-processing equipment that isn't adequately cleaned.

Biofilm infections develop gradually; sometimes, they do not cause symptoms immediately. They are rarely resolved by host defense mechanisms. Once an infection by a biofilm is established, it is very difficult to eradicate, because biofilms tend to be resistant to most of the methods used to control microbial growth, including antibiotics. Biofilms respond poorly or only temporarily to antibiotics; it has been said that they can resist up to 1,000 times the antibiotic concentrations used to kill the same bacteria when they are free-living or planktonic. An antibiotic dose that large would harm the patient; therefore, scientists are working on new ways to get rid of biofilms.

Antibiotics: Are We Facing a Crisis?

The word *antibiotic* comes from the Greek *anti* meaning "against" and *bios* meaning "life." An **antibiotic** is a chemical, produced either by microbes or synthetically, that is hostile to the growth of other organisms. Today's news and media often address concerns about an antibiotic crisis. Are the antibiotics that easily treated bacterial infections in the past becoming obsolete? Are there new "superbugs"—bacteria that have evolved to become more resistant to our arsenal of antibiotics? Is this the beginning of the end of antibiotics? All these questions challenge the healthcare community.

One of the main causes of resistant bacteria is the abuse of antibiotics. The imprudent and excessive use of antibiotics has resulted in the natural selection of resistant forms of bacteria. The antibiotic kills most of the infecting bacteria, and therefore only the resistant forms remain. These resistant forms reproduce, resulting in an increase in the proportion of resistant forms over non-resistant ones. Another major misuse of antibiotics is in patients with colds or the flu, for which antibiotics are useless. Another

problem is the excessive use of antibiotics in livestock. The routine use of antibiotics in animal feed promotes bacterial resistance as well. In the United States, 70 percent of the antibiotics produced are fed to animals. These antibiotics are given to livestock in low doses, which maximize the probability of resistance developing, and these resistant bacteria are readily transferred to humans.

Note:
Link to Learning

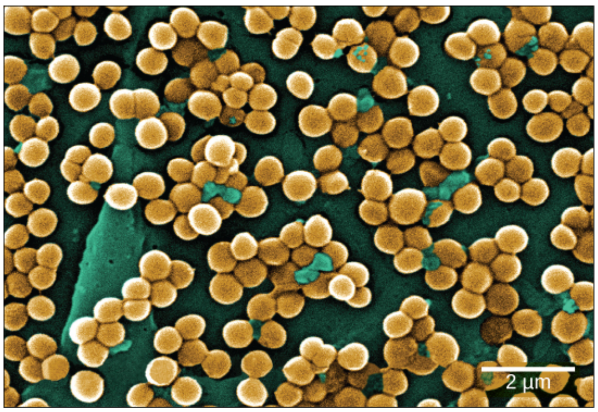
Watch a recent news report on the problem of routine antibiotic administration to livestock and antibiotic-resistant bacteria.

One of the Superbugs: MRSA

The imprudent use of antibiotics has paved the way for bacteria to expand populations of resistant forms. For example, *Staphylococcus aureu*s, often called "staph," is a common bacterium that can live in the human body and is usually easily treated with antibiotics. A very dangerous strain, however, **methicillin-resistant *Staphylococcus aureus* (MRSA)** has made the news over the past few years ([link]). This strain is resistant to many commonly used antibiotics, including methicillin, amoxicillin, penicillin, and oxacillin. MRSA can cause infections of the skin, but it can also infect the bloodstream, lungs, urinary tract, or sites of injury. While MRSA infections are common among people in healthcare facilities, they have also appeared in healthy people who haven't been hospitalized but who live or work in tight populations (like military personnel and prisoners). Researchers have expressed concern about the way this latter source of MRSA targets a much

younger population than those residing in care facilities. *The Journal of the American Medical Association* reported that, among MRSA-afflicted persons in healthcare facilities, the average age is 68, whereas people with "community-associated MRSA" (**CA-MRSA**) have an average age of 23. [footnote]

Naimi, TS, LeDell, KH, Como-Sabetti, K, et al. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. *JAMA* 290 (2003): 2976–84, doi: 10.1001/jama.290.22.2976.



This scanning electron micrograph shows methicillin-resistant *Staphylococcus aureus* bacteria, commonly known as MRSA. *S. aureus* is not always pathogenic, but can cause diseases such as food poisoning and skin and respiratory infections. (credit: modification of work by Janice Haney Carr; scale-bar data from Matt Russell)

In summary, the medical community is facing an antibiotic crisis. Some scientists believe that after years of being protected from bacterial

infections by antibiotics, we may be returning to a time in which a simple bacterial infection could again devastate the human population. Researchers are developing new antibiotics, but it takes many years to of research and clinical trials, plus financial investments in the millions of dollars, to generate an effective and approved drug.

Foodborne Diseases

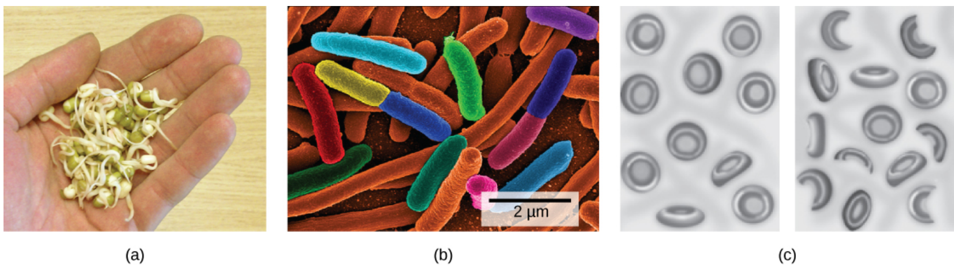
Prokaryotes are everywhere: They readily colonize the surface of any type of material, and food is not an exception. Most of the time, prokaryotes colonize food and food-processing equipment in the form of a biofilm. Outbreaks of bacterial infection related to food consumption are common. A **foodborne disease** (colloquially called “food poisoning”) is an illness resulting from the consumption the pathogenic bacteria, viruses, or other parasites that contaminate food. Although the United States has one of the safest food supplies in the world, the U.S. Centers for Disease Control and Prevention (CDC) has reported that “76 million people get sick, more than 300,000 are hospitalized, and 5,000 Americans die each year from foodborne illness.”

The characteristics of foodborne illnesses have changed over time. In the past, it was relatively common to hear about sporadic cases of **botulism**, the potentially fatal disease produced by a toxin from the anaerobic bacterium *Clostridium botulinum*. Some of the most common sources for this bacterium were non-acidic canned foods, homemade pickles, and processed meat and sausages. The can, jar, or package created a suitable anaerobic environment where *Clostridium* could grow. Proper sterilization and canning procedures have reduced the incidence of this disease.

While people may tend to think of foodborne illnesses as associated with animal-based foods, most cases are now linked to produce. There have been serious, produce-related outbreaks associated with raw spinach in the United States and with vegetable sprouts in Germany, and these types of outbreaks have become more common. The raw spinach outbreak in 2006 was produced by the bacterium *E. coli* serotype O157:H7. A **serotype** is a strain of bacteria that carries a set of similar antigens on its cell surface, and there are often many different serotypes of a bacterial species. Most *E. coli*

are not particularly dangerous to humans, but serotype O157:H7 can cause bloody diarrhea and is potentially fatal.

All types of food can potentially be contaminated with bacteria. Recent outbreaks of *Salmonella* reported by the CDC occurred in foods as diverse as peanut butter, alfalfa sprouts, and eggs. A deadly outbreak in Germany in 2010 was caused by *E. coli* contamination of vegetable sprouts ([link]). The strain that caused the outbreak was found to be a new serotype not previously involved in other outbreaks, which indicates that *E. coli* is continuously evolving.



(a) Vegetable sprouts grown at an organic farm were the cause of an (b) *E. coli* outbreak that killed 32 people and sickened 3,800 in Germany in 2011. The strain responsible, *E. coli* O104:H4, produces Shiga toxin, a substance that inhibits protein synthesis in the host cell. The toxin (c) destroys red blood cells resulting in bloody diarrhea. Deformed red blood cells clog the capillaries of the kidney, which can lead to kidney failure, as happened to 845 patients in the 2011 outbreak. Kidney failure is usually reversible, but some patients experience kidney problems years later. (credit c: NIDDK, NIH)

Note:

Career Connection

Epidemiologist

Epidemiology is the study of the occurrence, distribution, and determinants of health and disease in a population. It is, therefore, part of public health. An epidemiologist studies the frequency and distribution of diseases within human populations and environments.

Epidemiologists collect data about a particular disease and track its spread to identify the original mode of transmission. They sometimes work in close collaboration with historians to try to understand the way a disease evolved geographically and over time, tracking the natural history of pathogens. They gather information from clinical records, patient interviews, surveillance, and any other available means. That information is used to develop strategies, such as vaccinations ([link]), and design public health policies to reduce the incidence of a disease or to prevent its spread. Epidemiologists also conduct rapid investigations in case of an outbreak to recommend immediate measures to control it.

An epidemiologist has a bachelor’s degree, plus a master’s degree in public health (MPH). Many epidemiologists are also physicians (and have an M.D.), or they have a Ph.D. in an associated field, such as biology or microbiology.

Vaccinations can slow the spread of communicable diseases. (credit: modification of work by Daniel Paquet)

Section Summary

Devastating diseases and plagues have been among us since early times. There are records about microbial diseases as far back as 3000 B.C. Infectious diseases remain among the leading causes of death worldwide. Emerging diseases are those rapidly increasing in incidence or geographic range. They can be new or re-emerging diseases (previously under control). Many emerging diseases affecting humans, such as brucellosis, are zoonoses. The WHO has identified a group of diseases whose re-emergence should be monitored: Those caused by bacteria include bubonic plague, diphtheria, and cholera.

Biofilms are considered responsible for diseases such as bacterial infections in patients with cystic fibrosis, Legionnaires' disease, and otitis media. They produce dental plaque; colonize catheters, prostheses, transcutaneous, and orthopedic devices; and infect contact lenses, open wounds, and burned tissue. Biofilms also produce foodborne diseases because they colonize the surfaces of food and food-processing equipment. Biofilms are resistant to most of the methods used to control microbial growth. The excessive use of antibiotics has resulted in a major global problem, since resistant forms of bacteria have been selected over time. A very dangerous strain, methicillin-resistant *Staphylococcus aureus* (MRSA), has wreaked havoc recently. Foodborne diseases result from the consumption of contaminated food, pathogenic bacteria, viruses, or parasites that contaminate food.

Review Questions

Exercise:

Problem:

A disease that is constantly present in a population is called _____.

a. pandemic
b. epidemic
c. endemic
d. re-emerging

Solution:

C

Exercise:

Problem: Which of the statements about biofilms is incorrect?

a. Biofilms are considered responsible for diseases such as cystic fibrosis.
b. Biofilms produce dental plaque, and colonize catheters and prostheses.
c. Biofilms colonize open wounds and burned tissue.
d. All statements are incorrect.

Solution:

D

Exercise:

Problem: Which of these statements is true?

a. An antibiotic is any substance produced by a organism that is antagonistic to the growth of prokaryotes.
b. An antibiotic is any substance produced by a prokaryote that is antagonistic to the growth of other viruses.
c. An antibiotic is any substance produced by a prokaryote that is antagonistic to the growth of eukaryotic cells.
d. An antibiotic is any substance produced by a prokaryote that prevents growth of the same prokaryote.

Solution:

A

Free Response

Exercise:

Problem:

Explain the reason why the imprudent and excessive use of antibiotics has resulted in a major global problem.

Solution:

Antibiotics kill bacteria that are sensitive to them; thus, only the resistant ones will survive. These resistant bacteria will reproduce, and therefore, after a while, there will be only resistant bacteria.

Exercise:

Problem:

Researchers have discovered that washing spinach with water several times does not prevent foodborne diseases due to *E. coli*. How can you explain this fact?

Solution:

E. coli colonizes the surface of the leaf, forming a biofilm that is more difficult to remove than free (planktonic) cells. Additionally, bacteria can be taken up in the water that plants are grown in, thereby entering the plant tissues rather than simply residing on the leaf surface.

Glossary

antibiotic
: biological substance that, in low concentration, is antagonistic to the growth of prokaryotes

Black Death
: devastating pandemic that is believed to have been an outbreak of bubonic plague caused by the bacterium *Yersinia pestis*

botulism
: disease produced by the toxin of the anaerobic bacterium *Clostridium botulinum*

CA-MRSA
: MRSA acquired in the community rather than in a hospital setting

emerging disease
: disease making an initial appearance in a population or that is increasing in incidence or geographic range

endemic disease
: disease that is constantly present, usually at low incidence, in a population

epidemic
: disease that occurs in an unusually high number of individuals in a population at the same time

foodborne disease
: any illness resulting from the consumption of contaminated food, or of the pathogenic bacteria, viruses, or other parasites that contaminate food

MRSA
: (methicillin-resistant *Staphylococcus aureus*) very dangerous *Staphylococcus aureus* strain resistant to multiple antibiotics

pandemic
: widespread, usually worldwide, epidemic disease

serotype
: strain of bacteria that carries a set of similar antigens on its cell surface, often many in a bacterial species

zoonosis
: disease that primarily infects animals that is transmitted to humans

Eukaryotic Origins
By the end of this section, you will be able to:

- List the unifying characteristics of eukaryotes
- Describe what scientists know about the origins of eukaryotes based on the last common ancestor
- Explain endosymbiotic theory

Living things fall into three large groups: Archaea, Bacteria, and Eukarya. The first two have prokaryotic cells, and the third contains all eukaryotes. A relatively sparse fossil record is available to help discern what the first members of each of these lineages looked like, so it is possible that all the events that led to the last common ancestor of extant eukaryotes will remain unknown. However, comparative biology of extant organisms and the limited fossil record provide some insight into the history of Eukarya.

The earliest fossils found appear to be Bacteria, most likely cyanobacteria. They are about 3.5 billion years old and are recognizable because of their relatively complex structure and, for prokaryotes, relatively large cells. Most other prokaryotes have small cells, 1 or 2 µm in size, and would be difficult to pick out as fossils. Most living eukaryotes have cells measuring 10 µm or greater. Structures this size, which might be fossils, appear in the geological record about 2.1 billion years ago.

Characteristics of Eukaryotes

Data from these fossils have led comparative biologists to the conclusion that living eukaryotes are all descendants of a single common ancestor. Mapping the characteristics found in all major groups of eukaryotes reveals that the following characteristics must have been present in the last common ancestor, because these characteristics are present in at least some of the members of each major lineage.

1. Cells with nuclei surrounded by a nuclear envelope with nuclear pores. This is the single characteristic that is both necessary and sufficient to define an organism as a eukaryote. All extant eukaryotes have cells with nuclei.

2. Mitochondria. Some extant eukaryotes have very reduced remnants of mitochondria in their cells, whereas other members of their lineages have “typical” mitochondria.
3. A cytoskeleton containing the structural and motility components called actin microfilaments and microtubules. All extant eukaryotes have these cytoskeletal elements.
4. Flagella and cilia, organelles associated with cell motility. Some extant eukaryotes lack flagella and/or cilia, but they are descended from ancestors that possessed them.
5. Chromosomes, each consisting of a linear DNA molecule coiled around basic (alkaline) proteins called histones. The few eukaryotes with chromosomes lacking histones clearly evolved from ancestors that had them.
6. Mitosis, a process of nuclear division wherein replicated chromosomes are divided and separated using elements of the cytoskeleton. Mitosis is universally present in eukaryotes.
7. Sex, a process of genetic recombination unique to eukaryotes in which diploid nuclei at one stage of the life cycle undergo meiosis to yield haploid nuclei and subsequent karyogamy, a stage where two haploid nuclei fuse together to create a diploid zygote nucleus.
8. Members of all major lineages have cell walls, and it might be reasonable to conclude that the last common ancestor could make cell walls during some stage of its life cycle. However, not enough is known about eukaryotes’ cell walls and their development to know how much homology exists among them. If the last common ancestor could make cell walls, it is clear that this ability must have been lost in many groups.

Endosymbiosis and the Evolution of Eukaryotes

In order to understand eukaryotic organisms fully, it is necessary to understand that all extant eukaryotes are descendants of a chimeric organism that was a composite of a host cell and the cell(s) of an alpha-proteobacterium that “took up residence” inside it. This major theme in the origin of eukaryotes is known as **endosymbiosis**, one cell engulfing another such that the engulfed cell survives and both cells benefit. Over many generations, a symbiotic relationship can result in two organisms that

depend on each other so completely that neither could survive on its own. Endosymbiotic events likely contributed to the origin of the last common ancestor of today’s eukaryotes and to later diversification in certain lineages of eukaryotes ([link]). Before explaining this further, it is necessary to consider metabolism in prokaryotes.

Prokaryotic Metabolism

Many important metabolic processes arose in prokaryotes, and some of these, such as nitrogen fixation, are never found in eukaryotes. The process of aerobic respiration is found in all major lineages of eukaryotes, and it is localized in the mitochondria. Aerobic respiration is also found in many lineages of prokaryotes, but it is not present in all of them, and many forms of evidence suggest that such anaerobic prokaryotes never carried out aerobic respiration nor did their ancestors.

While today’s atmosphere is about one-fifth molecular oxygen (O_2), geological evidence shows that it originally lacked O_2. Without oxygen, aerobic respiration would not be expected, and living things would have relied on fermentation instead. At some point before, about 3.5 billion years ago, some prokaryotes began using energy from sunlight to power anabolic processes that reduce carbon dioxide to form organic compounds. That is, they evolved the ability to photosynthesize. Hydrogen, derived from various sources, was captured using light-powered reactions to reduce fixed carbon dioxide in the Calvin cycle. The group of Gram-negative bacteria that gave rise to cyanobacteria used water as the hydrogen source and released O_2 as a waste product.

Eventually, the amount of photosynthetic oxygen built up in some environments to levels that posed a risk to living organisms, since it can damage many organic compounds. Various metabolic processes evolved that protected organisms from oxygen, one of which, aerobic respiration, also generated high levels of ATP. It became widely present among prokaryotes, including in a group we now call alpha-proteobacteria. Organisms that did not acquire aerobic respiration had to remain in oxygen-free environments. Originally, oxygen-rich environments were likely

localized around places where cyanobacteria were active, but by about 2 billion years ago, geological evidence shows that oxygen was building up to higher concentrations in the atmosphere. Oxygen levels similar to today's levels only arose within the last 700 million years.

Recall that the first fossils that we believe to be eukaryotes date to about 2 billion years old, so they appeared as oxygen levels were increasing. Also, recall that all extant eukaryotes descended from an ancestor with mitochondria. These organelles were first observed by light microscopists in the late 1800s, where they appeared to be somewhat worm-shaped structures that seemed to be moving around in the cell. Some early observers suggested that they might be bacteria living inside host cells, but these hypotheses remained unknown or rejected in most scientific communities.

Endosymbiotic Theory

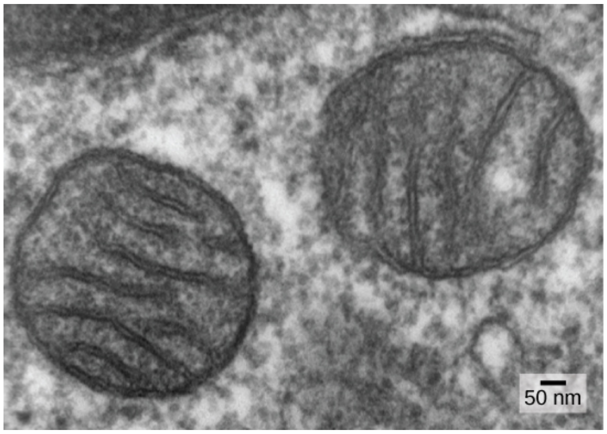
As cell biology developed in the twentieth century, it became clear that mitochondria were the organelles responsible for producing ATP using aerobic respiration. In the 1960s, American biologist Lynn Margulis developed **endosymbiotic theory**, which states that eukaryotes may have been a product of one cell engulfing another, one living within another, and evolving over time until the separate cells were no longer recognizable as such. In 1967, Margulis introduced new work on the theory and substantiated her findings through microbiological evidence. Although Margulis' work initially was met with resistance, this once-revolutionary hypothesis is now widely (but not completely) accepted, with work progressing on uncovering the steps involved in this evolutionary process and the key players involved. Much still remains to be discovered about the origins of the cells that now make up the cells in all living eukaryotes.

Broadly, it has become clear that many of our nuclear genes and the molecular machinery responsible for replication and expression appear closely related to those in Archaea. On the other hand, the metabolic organelles and genes responsible for many energy-harvesting processes had their origins in bacteria. Much remains to be clarified about how this

relationship occurred; this continues to be an exciting field of discovery in biology. For instance, it is not known whether the endosymbiotic event that led to mitochondria occurred before or after the host cell had a nucleus. Such organisms would be among the extinct precursors of the last common ancestor of eukaryotes.

Mitochondria

One of the major features distinguishing prokaryotes from eukaryotes is the presence of mitochondria. Eukaryotic cells may contain anywhere from one to several thousand mitochondria, depending on the cell's level of energy consumption. Each mitochondrion measures 1 to 10 or greater micrometers in length and exists in the cell as an organelle that can be ovoid to worm-shaped to intricately branched ([link]). Mitochondria arise from the division of existing mitochondria; they may fuse together; and they may be moved around inside the cell by interactions with the cytoskeleton. However, mitochondria cannot survive outside the cell. As the atmosphere was oxygenated by photosynthesis, and as successful aerobic prokaryotes evolved, evidence suggests that an ancestral cell with some membrane compartmentalization engulfed a free-living aerobic prokaryote, specifically an alpha-proteobacterium, thereby giving the host cell the ability to use oxygen to release energy stored in nutrients. Alpha-proteobacteria are a large group of bacteria that includes species symbiotic with plants, disease organisms that can infect humans via ticks, and many free-living species that use light for energy. Several lines of evidence support that mitochondria are derived from this endosymbiotic event. Most mitochondria are shaped like alpha-proteobacteria and are surrounded by two membranes. The mitochondrial inner membrane is extensive and involves substantial infoldings called cristae that resemble the textured, outer surface of alpha-proteobacteria. The matrix and inner membrane are rich with the enzymes necessary for aerobic respiration.



In this transmission electron micrograph of mitochondria in a mammalian lung cell, the cristae, infoldings of the mitochondrial inner membrane, can be seen in cross-section. (credit: Louise Howard)

Mitochondria divide independently by a process that resembles binary fission in prokaryotes. Specifically, mitochondria are not formed from scratch (de novo) by the eukaryotic cell; they reproduce within it and are distributed with the cytoplasm when a cell divides or two cells fuse. Therefore, although these organelles are highly integrated into the eukaryotic cell, they still reproduce as if they are independent organisms within the cell. However, their reproduction is synchronized with the activity and division of the cell. Mitochondria have their own (usually) circular DNA chromosome that is stabilized by attachments to the inner membrane and carries genes similar to genes expressed by alpha-proteobacteria. Mitochondria also have special ribosomes and transfer RNAs that resemble these components in prokaryotes. These features all support that mitochondria were once free-living prokaryotes.

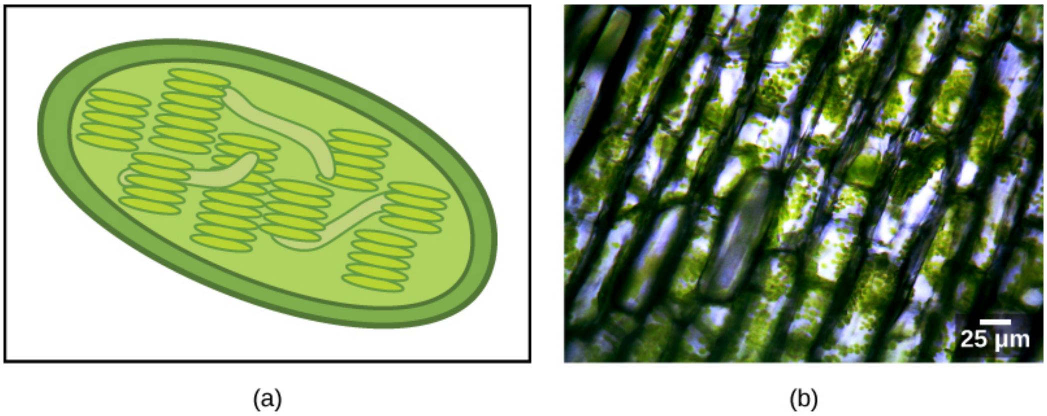
Mitochondria that carry out aerobic respiration have their own genomes, with genes similar to those in alpha-proteobacteria. However, many of the genes for respiratory proteins are located in the nucleus. When these genes

are compared to those of other organisms, they appear to be of alpha-proteobacterial origin. Additionally, in some eukaryotic groups, such genes are found in the mitochondria, whereas in other groups, they are found in the nucleus. This has been interpreted as evidence that genes have been transferred from the endosymbiont chromosome to the host genome. This loss of genes by the endosymbiont is probably one explanation why mitochondria cannot live without a host.

Some living eukaryotes are anaerobic and cannot survive in the presence of too much oxygen. Some appear to lack organelles that could be recognized as mitochondria. In the 1970s to the early 1990s, many biologists suggested that some of these eukaryotes were descended from ancestors whose lineages had diverged from the lineage of mitochondrion-containing eukaryotes before endosymbiosis occurred. However, later findings suggest that reduced organelles are found in most, if not all, anaerobic eukaryotes, and that all eukaryotes appear to carry some genes in their nuclei that are of mitochondrial origin. In addition to the aerobic generation of ATP, mitochondria have several other metabolic functions. One of these functions is to generate clusters of iron and sulfur that are important cofactors of many enzymes. Such functions are often associated with the reduced mitochondrion-derived organelles of anaerobic eukaryotes. Therefore, most biologists accept that the last common ancestor of eukaryotes had mitochondria.

Plastids

Some groups of eukaryotes are photosynthetic. Their cells contain, in addition to the standard eukaryotic organelles, another kind of organelle called a **plastid**. When such cells are carrying out photosynthesis, their plastids are rich in the pigment chlorophyll *a* and a range of other pigments, called accessory pigments, which are involved in harvesting energy from light. Photosynthetic plastids are called chloroplasts ([link]).



(a) This chloroplast cross-section illustrates its elaborate inner membrane organization. Stacks of thylakoid membranes compartmentalize photosynthetic enzymes and provide scaffolding for chloroplast DNA. (b) In this micrograph of *Elodea* sp., the chloroplasts can be seen as small green spheres. (credit b: modification of work by Brandon Zierer; scale-bar data from Matt Russell)

Like mitochondria, plastids appear to have an endosymbiotic origin. This hypothesis was also championed by Lynn Margulis. Plastids are derived from cyanobacteria that lived inside the cells of an ancestral, aerobic, heterotrophic eukaryote. This is called primary endosymbiosis, and plastids of primary origin are surrounded by two membranes. The best evidence is that this has happened twice in the history of eukaryotes. In one case, the common ancestor of the major lineage/supergroup Archaeplastida took on a cyanobacterial endosymbiont; in the other, the ancestor of the small amoeboid rhizarian taxon, *Paulinella*, took on a different cyanobacterial endosymbiont. Almost all photosynthetic eukaryotes are descended from the first event, and only a couple of species are derived from the other.

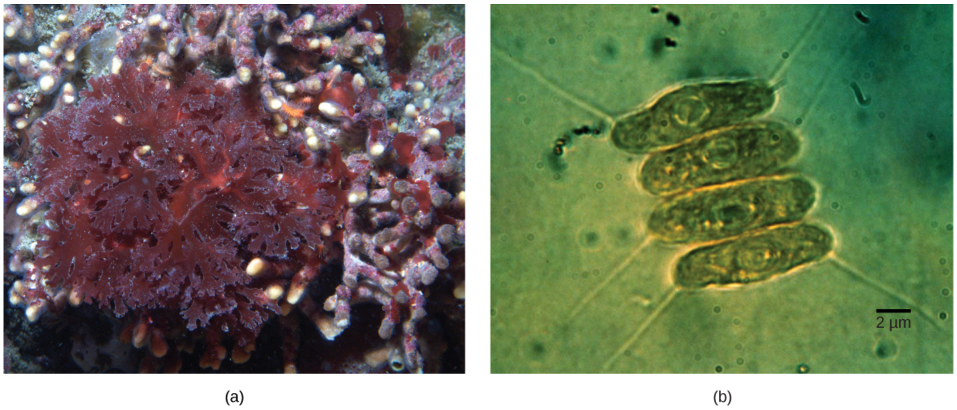
Cyanobacteria are a group of Gram-negative bacteria with all the conventional structures of the group. However, unlike most prokaryotes, they have extensive, internal membrane-bound sacs called thylakoids. Chlorophyll is a component of these membranes, as are many of the proteins of the light reactions of photosynthesis. Cyanobacteria also have

the peptidoglycan wall and lipopolysaccharide layer associated with Gram-negative bacteria.

Chloroplasts of primary origin have thylakoids, a circular DNA chromosome, and ribosomes similar to those of cyanobacteria. Each chloroplast is surrounded by two membranes. In the group of Archaeplastida called the glaucophytes and in *Paulinella*, a thin peptidoglycan layer is present between the outer and inner plastid membranes. All other plastids lack this relictual cyanobacterial wall. The outer membrane surrounding the plastid is thought to be derived from the vacuole in the host, and the inner membrane is thought to be derived from the plasma membrane of the symbiont.

There is also, as with the case of mitochondria, strong evidence that many of the genes of the endosymbiont were transferred to the nucleus. Plastids, like mitochondria, cannot live independently outside the host. In addition, like mitochondria, plastids are derived from the division of other plastids and never built from scratch. Researchers have suggested that the endosymbiotic event that led to Archaeplastida occurred 1 to 1.5 billion years ago, at least 5 hundred million years after the fossil record suggests that eukaryotes were present.

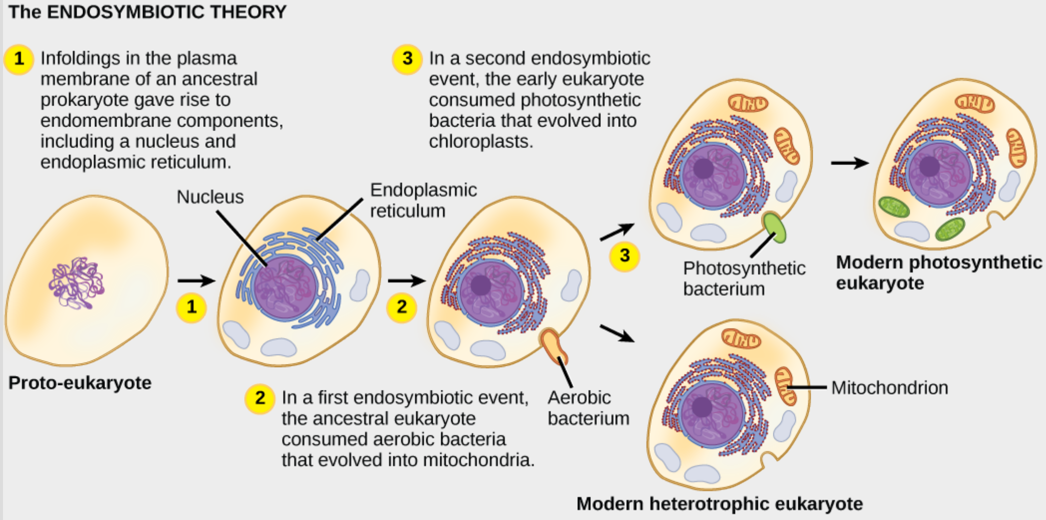
Not all plastids in eukaryotes are derived directly from primary endosymbiosis. Some of the major groups of algae became photosynthetic by secondary endosymbiosis, that is, by taking in either green algae or red algae (both from Archaeplastida) as endosymbionts ([link]**ab**). Numerous microscopic and genetic studies have supported this conclusion. Secondary plastids are surrounded by three or more membranes, and some secondary plastids even have clear remnants of the nucleus of endosymbiotic alga. Others have not “kept” any remnants. There are cases where tertiary or higher-order endosymbiotic events are the best explanations for plastids in some eukaryotes.



(a) Red algae and (b) green algae (visualized by light microscopy) share similar DNA sequences with photosynthetic cyanobacteria. Scientists speculate that, in a process called endosymbiosis, an ancestral prokaryote engulfed a photosynthetic cyanobacterium that evolved into modern-day chloroplasts. (credit a: modification of work by Ed Bierman; credit b: modification of work by G. Fahnenstiel, NOAA; scale-bar data from Matt Russell)

Note:

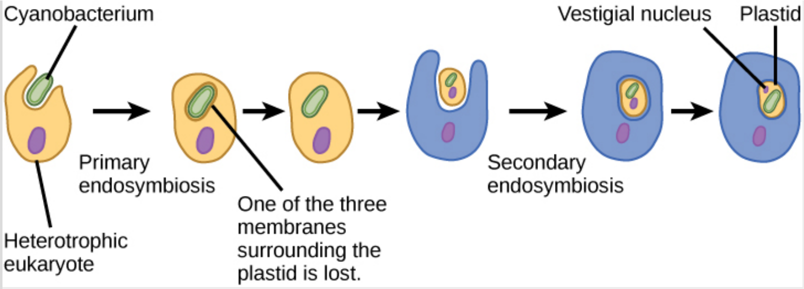
Art Connection



The first eukaryote may have originated from an ancestral prokaryote that had undergone membrane proliferation, compartmentalization of cellular function (into a nucleus, lysosomes, and an endoplasmic reticulum), and the establishment of endosymbiotic relationships with an aerobic prokaryote, and, in some cases, a photosynthetic prokaryote, to form mitochondria and chloroplasts, respectively.

What evidence is there that mitochondria were incorporated into the ancestral eukaryotic cell before chloroplasts?

Note:
Evolution Connection
Secondary Endosymbiosis in Chlorarachniophytes
Endosymbiosis involves one cell engulfing another to produce, over time, a coevolved relationship in which neither cell could survive alone. The chloroplasts of red and green algae, for instance, are derived from the engulfment of a photosynthetic cyanobacterium by an early prokaryote. This leads to the question of the possibility of a cell containing an endosymbiont to itself become engulfed, resulting in a secondary endosymbiosis. Molecular and morphological evidence suggest that the chlorarachniophyte protists are derived from a secondary endosymbiotic event. Chlorarachniophytes are rare algae indigenous to tropical seas and sand that can be classified into the rhizarian supergroup. Chlorarachniophytes extend thin cytoplasmic strands, interconnecting themselves with other chlorarachniophytes, in a cytoplasmic network. These protists are thought to have originated when a eukaryote engulfed a green alga, the latter of which had already established an endosymbiotic relationship with a photosynthetic cyanobacterium ([link]).



The hypothesized process of endosymbiotic events leading to the evolution of chlorarachniophytes is shown. In a primary endosymbiotic event, a heterotrophic eukaryote consumed a cyanobacterium. In a secondary endosymbiotic event, the cell resulting from primary endosymbiosis was consumed by a second cell. The resulting organelle became a plastid in modern chlorarachniophytes.

Several lines of evidence support that chlorarachniophytes evolved from secondary endosymbiosis. The chloroplasts contained within the green algal endosymbionts still are capable of photosynthesis, making chlorarachniophytes photosynthetic. The green algal endosymbiont also exhibits a stunted vestigial nucleus. In fact, it appears that chlorarachniophytes are the products of an evolutionarily recent secondary endosymbiotic event. The plastids of chlorarachniophytes are surrounded by four membranes: The first two correspond to the inner and outer membranes of the photosynthetic cyanobacterium, the third corresponds to the green alga, and the fourth corresponds to the vacuole that surrounded the green alga when it was engulfed by the chlorarachniophyte ancestor. In other lineages that involved secondary endosymbiosis, only three membranes can be identified around plastids. This is currently rectified as a sequential loss of a membrane during the course of evolution.

The process of secondary endosymbiosis is not unique to chlorarachniophytes. In fact, secondary endosymbiosis of green algae also led to euglenid protists, whereas secondary endosymbiosis of red algae led to the evolution of dinoflagellates, apicomplexans, and stramenopiles.

Section Summary

The oldest fossil evidence of eukaryotes is about 2 billion years old. Fossils older than this all appear to be prokaryotes. It is probable that today's eukaryotes are descended from an ancestor that had a prokaryotic organization. The last common ancestor of today's Eukarya had several characteristics, including cells with nuclei that divided mitotically and contained linear chromosomes where the DNA was associated with histones, a cytoskeleton and endomembrane system, and the ability to make cilia/flagella during at least part of its life cycle. It was aerobic because it had mitochondria that were the result of an aerobic alpha-proteobacterium that lived inside a host cell. Whether this host had a nucleus at the time of the initial symbiosis remains unknown. The last common ancestor may have had a cell wall for at least part of its life cycle, but more data are needed to confirm this hypothesis. Today's eukaryotes are very diverse in their shapes, organization, life cycles, and number of cells per individual.

Art Connections

Exercise:

Problem:

[link] What evidence is there that mitochondria were incorporated into the ancestral eukaryotic cell before chloroplasts?

Solution:

[link] All eukaryotic cells have mitochondria, but not all eukaryotic cells have chloroplasts.

Review Questions

Exercise:

Problem:

What event is thought to have contributed to the evolution of eukaryotes?

a. global warming
b. glaciation
c. volcanic activity
d. oxygenation of the atmosphere

Solution:

D

Exercise:

Problem:

Which characteristic is shared by prokaryotes and eukaryotes?

a. cytoskeleton
b. nuclear envelope
c. DNA-based genome
d. mitochondria

Solution:

C

Exercise:

Problem: Mitochondria most likely evolved by ______________.

a. a photosynthetic cyanobacterium
b. cytoskeletal elements
c. endosymbiosis
d. membrane proliferation

Solution:

C

Exercise:

Problem:

Which of these protists is believed to have evolved following a secondary endosymbiosis?

a. green algae
b. cyanobacteria
c. red algae
d. chlorarachniophytes

Solution:

D

Free Response

Exercise:

Problem:

Describe the hypothesized steps in the origin of eukaryotic cells.

Solution:

Eukaryotic cells arose through endosymbiotic events that gave rise to the energy-producing organelles within the eukaryotic cells such as mitochondria and chloroplasts. The nuclear genome of eukaryotes is related most closely to the Archaea, so it may have been an early archaean that engulfed a bacterial cell that evolved into a mitochondrion. Mitochondria appear to have originated from an alpha-proteobacterium, whereas chloroplasts originated as a cyanobacterium.

There is also evidence of secondary endosymbiotic events. Other cell components may also have resulted from endosymbiotic events.

Glossary

endosymbiosis
: engulfment of one cell within another such that the engulfed cell survives, and both cells benefit; the process responsible for the evolution of mitochondria and chloroplasts in eukaryotes

endosymbiotic theory
: theory that states that eukaryotes may have been a product of one cell engulfing another, one living within another, and evolving over time until the separate cells were no longer recognizable as such

plastid
: one of a group of related organelles in plant cells that are involved in the storage of starches, fats, proteins, and pigments

Characteristics of Protists
By the end of this section, you will be able to:

- Describe the cell structure characteristics of protists
- Describe the metabolic diversity of protists
- Describe the life cycle diversity of protists

There are over 100,000 described living species of protists, and it is unclear how many undescribed species may exist. Since many protists live as commensals or parasites in other organisms and these relationships are often species-specific, there is a huge potential for protist diversity that matches the diversity of hosts. As the catchall term for eukaryotic organisms that are not animal, plant, or fungi, it is not surprising that very few characteristics are common to all protists.

Cell Structure

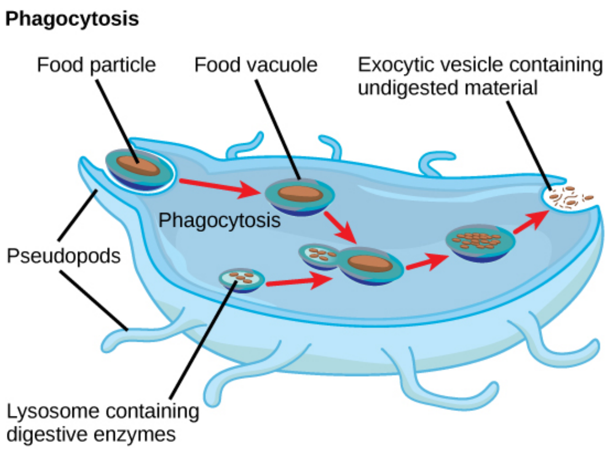
The cells of protists are among the most elaborate of all cells. Most protists are microscopic and unicellular, but some true multicellular forms exist. A few protists live as colonies that behave in some ways as a group of free-living cells and in other ways as a multicellular organism. Still other protists are composed of enormous, multinucleate, single cells that look like amorphous blobs of slime, or in other cases, like ferns. In fact, many protist cells are multinucleated; in some species, the nuclei are different sizes and have distinct roles in protist cell function.

Single protist cells range in size from less than a micrometer to three meters in length to hectares. Protist cells may be enveloped by animal-like cell membranes or plant-like cell walls. Others are encased in glassy silica-based shells or wound with **pellicles** of interlocking protein strips. The pellicle functions like a flexible coat of armor, preventing the protist from being torn or pierced without compromising its range of motion.

Metabolism

Protists exhibit many forms of nutrition and may be aerobic or anaerobic. Protists that store energy by photosynthesis belong to a group of

photoautotrophs and are characterized by the presence of chloroplasts. Other protists are heterotrophic and consume organic materials (such as other organisms) to obtain nutrition. Amoebas and some other heterotrophic protist species ingest particles by a process called phagocytosis, in which the cell membrane engulfs a food particle and brings it inward, pinching off an intracellular membranous sac, or vesicle, called a food vacuole ([link]). The vesicle containing the ingested particle, the phagosome, then fuses with a lysosome containing hydrolytic enzymes to produce a **phagolysosome**, and the food particle is broken down into small molecules that can diffuse into the cytoplasm and be used in cellular metabolism. Undigested remains ultimately are expelled from the cell via exocytosis.



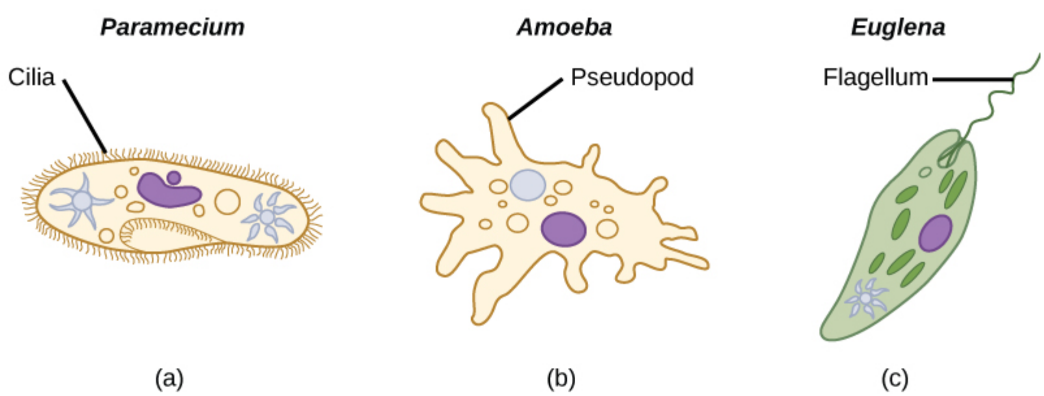
The stages of phagocytosis include the engulfment of a food particle, the digestion of the particle using hydrolytic enzymes contained within a lysosome, and the expulsion of undigested materials from the cell.

Subtypes of heterotrophs, called saprobes, absorb nutrients from dead organisms or their organic wastes. Some protists can function as

mixotrophs, obtaining nutrition by photoautotrophic or heterotrophic routes, depending on whether sunlight or organic nutrients are available.

Motility

The majority of protists are motile, but different types of protists have evolved varied modes of movement ([link]). Some protists have one or more flagella, which they rotate or whip. Others are covered in rows or tufts of tiny cilia that they coordinately beat to swim. Still others form cytoplasmic extensions called pseudopodia anywhere on the cell, anchor the pseudopodia to a substrate, and pull themselves forward. Some protists can move toward or away from a stimulus, a movement referred to as taxis. Movement toward light, termed phototaxis, is accomplished by coupling their locomotion strategy with a light-sensing organ.



Protists use various methods for transportation. (a) *Paramecium* waves hair-like appendages called cilia to propel itself. (b) *Amoeba* uses lobe-like pseudopodia to anchor itself to a solid surface and pull itself forward. (c) *Euglena* uses a whip-like tail called a flagellum to propel itself.

Life Cycles

Protists reproduce by a variety of mechanisms. Most undergo some form of asexual reproduction, such as binary fission, to produce two daughter cells. In protists, binary fission can be divided into transverse or longitudinal, depending on the axis of orientation; sometimes *Paramecium* exhibits this method. Some protists such as the true slime molds exhibit multiple fission and simultaneously divide into many daughter cells. Others produce tiny buds that go on to divide and grow to the size of the parental protist. Sexual reproduction, involving meiosis and fertilization, is common among protists, and many protist species can switch from asexual to sexual reproduction when necessary. Sexual reproduction is often associated with periods when nutrients are depleted or environmental changes occur. Sexual reproduction may allow the protist to recombine genes and produce new variations of progeny that may be better suited to surviving in the new environment. However, sexual reproduction is often associated with resistant cysts that are a protective, resting stage. Depending on their habitat, the cysts may be particularly resistant to temperature extremes, desiccation, or low pH. This strategy also allows certain protists to “wait out” stressors until their environment becomes more favorable for survival or until they are carried (such as by wind, water, or transport on a larger organism) to a different environment, because cysts exhibit virtually no cellular metabolism.

Protist life cycles range from simple to extremely elaborate. Certain parasitic protists have complicated life cycles and must infect different host species at different developmental stages to complete their life cycle. Some protists are unicellular in the haploid form and multicellular in the diploid form, a strategy employed by animals. Other protists have multicellular stages in both haploid and diploid forms, a strategy called alternation of generations that is also used by plants.

Habitats

Nearly all protists exist in some type of aquatic environment, including freshwater and marine environments, damp soil, and even snow. Several protist species are parasites that infect animals or plants. A few protist species live on dead organisms or their wastes, and contribute to their decay.

Section Summary

Protists are extremely diverse in terms of their biological and ecological characteristics, partly because they are an artificial assemblage of phylogenetically unrelated groups. Protists display highly varied cell structures, several types of reproductive strategies, virtually every possible type of nutrition, and varied habitats. Most single-celled protists are motile, but these organisms use diverse structures for transportation.

Review Questions

Exercise:

Problem:

Protists that have a pellicle are surrounded by _______________.

a. silica dioxide
b. calcium carbonate
c. carbohydrates
d. proteins

Solution:

D

Exercise:

Problem:

Protists with the capabilities to perform photosynthesis and to absorb nutrients from dead organisms are called _______________.

a. photoautotrophs
b. mixotrophs
c. saprobes
d. heterotrophs

Solution:

B

Exercise:

Problem:

Which of these locomotor organs would likely be the shortest?

a. a flagellum
b. a cilium
c. an extended pseudopod
d. a pellicle

Solution:

B

Exercise:

Problem: Alternation of generations describes which of the following?

a. The haploid form can be multicellular; the diploid form is unicellular.
b. The haploid form is unicellular; the diploid form can be multicellular.
c. Both the haploid and diploid forms can be multicellular.
d. Neither the haploid nor the diploid forms can be multicellular.

Solution:

C

Free Response

Exercise:

Problem:

Explain in your own words why sexual reproduction can be useful if a protist's environment changes.

Solution:

The ability to perform sexual reproduction allows protists to recombine their genes and produce new variations of progeny that may be better suited to the new environment. In contrast, asexual reproduction generates progeny that are clones of the parent.

Exercise:

Problem:

Giardia lamblia is a cyst-forming protist parasite that causes diarrhea if ingested. Given this information, against what type(s) of environments might *G. lamblia* cysts be particularly resistant?

Solution:

As an intestinal parasite, *Giardia* cysts would be exposed to low pH in the stomach acids of its host. To survive this environment and reach the intestine, the cysts would have to be resistant to acidic conditions.

Glossary

mixotroph
: organism that can obtain nutrition by autotrophic or heterotrophic means, usually facultatively

pellicle
: outer cell covering composed of interlocking protein strips that function like a flexible coat of armor, preventing cells from being torn or pierced without compromising their range of motion

phagolysosome

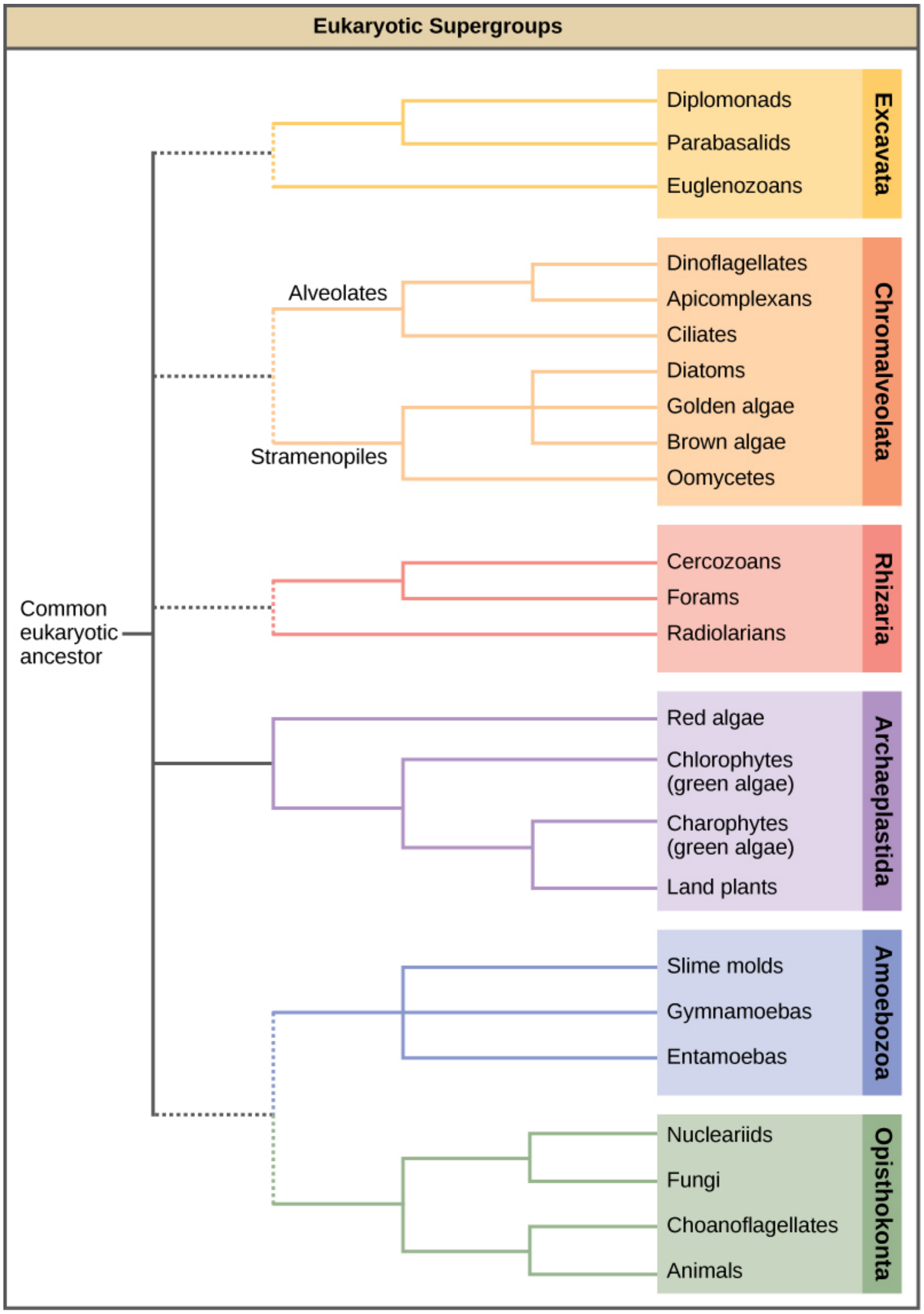
cellular body formed by the union of a phagosome containing the ingested particle with a lysosome that contains hydrolytic enzymes

Groups of Protists

By the end of this section, you will be able to:

- Describe representative protist organisms from each of the six presently recognized supergroups of eukaryotes
- Identify the evolutionary relationships of plants, animals, and fungi within the six presently recognized supergroups of eukaryotes

In the span of several decades, the Kingdom Protista has been disassembled because sequence analyses have revealed new genetic (and therefore evolutionary) relationships among these eukaryotes. Moreover, protists that exhibit similar morphological features may have evolved analogous structures because of similar selective pressures—rather than because of recent common ancestry. This phenomenon, called convergent evolution, is one reason why protist classification is so challenging. The emerging classification scheme groups the entire domain Eukaryota into six "supergroups" that contain all of the protists as well as animals, plants, and fungi that evolved from a common ancestor ([link]). The supergroups are believed to be monophyletic, meaning that all organisms within each supergroup are believed to have evolved from a single common ancestor, and thus all members are most closely related to each other than to organisms outside that group. There is still evidence lacking for the monophyly of some groups.



This diagram shows a proposed classification of the domain Eukara. Currently, the domain Eukarya is divided into six supergroups. Within each supergroup are multiple kingdoms. Dotted lines indicate suggested evolutionary relationships that remain under debate.

The classification of eukaryotes is still in flux, and the six supergroups may be modified or replaced by a more appropriate hierarchy as genetic, morphological, and ecological data accumulate. Keep in mind that the classification scheme presented here is just one of several hypotheses, and the true evolutionary relationships are still to be determined. When learning about protists, it is helpful to focus less on the nomenclature and more on the commonalities and differences that define the groups themselves.

Excavata

Many of the protist species classified into the supergroup Excavata are asymmetrical, single-celled organisms with a feeding groove “excavated” from one side. This supergroup includes heterotrophic predators, photosynthetic species, and parasites. Its subgroups are the diplomonads, parabasalids, and euglenozoans.

Diplomonads

Among the Excavata are the diplomonads, which include the intestinal parasite, *Giardia lamblia* ([link]). Until recently, these protists were believed to lack mitochondria. Mitochondrial remnant organelles, called **mitosomes**, have since been identified in diplomonads, but these mitosomes are essentially nonfunctional. Diplomonads exist in anaerobic environments and use alternative pathways, such as glycolysis, to generate energy. Each diplomonad cell has two identical nuclei and uses several flagella for locomotion.

The mammalian intestinal parasite *Giardia lamblia*, visualized here using scanning electron microscopy, is a waterborne protist that causes severe diarrhea when ingested. (credit: modification of work by Janice Carr, CDC; scale-bar data from Matt Russell)

Parabasalids

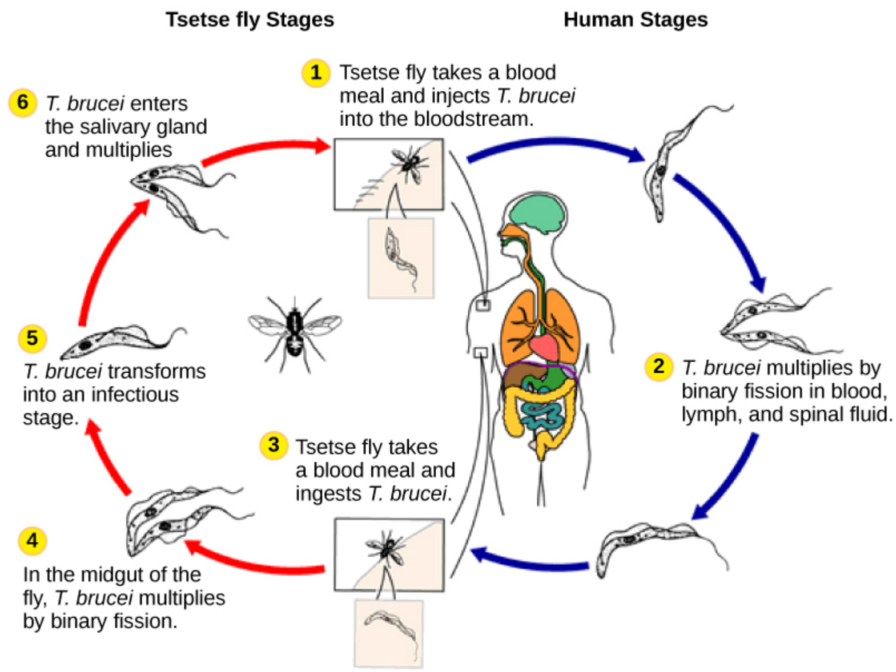
A second Excavata subgroup, the parabasalids, also exhibits semi-functional mitochondria. In parabasalids, these structures function anaerobically and are called **hydrogenosomes** because they produce hydrogen gas as a byproduct. Parabasalids move with flagella and membrane rippling. *Trichomonas vaginalis*, a parabasalid that causes a sexually transmitted disease in humans, employs these mechanisms to transit through the male and female urogenital tracts. *T. vaginalis* causes

trichamoniasis, which appears in an estimated 180 million cases worldwide each year. Whereas men rarely exhibit symptoms during an infection with this protist, infected women may become more susceptible to secondary infection with human immunodeficiency virus (HIV) and may be more likely to develop cervical cancer. Pregnant women infected with *T. vaginalis* are at an increased risk of serious complications, such as pre-term delivery.

Euglenozoans

Euglenozoans includes parasites, heterotrophs, autotrophs, and mixotrophs, ranging in size from 10 to 500 µm. Euglenoids move through their aquatic habitats using two long flagella that guide them toward light sources sensed by a primitive ocular organ called an eyespot. The familiar genus, *Euglena*, encompasses some mixotrophic species that display a photosynthetic capability only when light is present. In the dark, the chloroplasts of *Euglena* shrink up and temporarily cease functioning, and the cells instead take up organic nutrients from their environment.

The human parasite, *Trypanosoma brucei*, belongs to a different subgroup of Euglenozoa, the kinetoplastids. The kinetoplastid subgroup is named after the **kinetoplast**, a DNA mass carried within the single, oversized mitochondrion possessed by each of these cells. This subgroup includes several parasites, collectively called trypanosomes, which cause devastating human diseases and infect an insect species during a portion of their life cycle. *T. brucei* develops in the gut of the tsetse fly after the fly bites an infected human or other mammalian host. The parasite then travels to the insect salivary glands to be transmitted to another human or other mammal when the infected tsetse fly consumes another blood meal. *T. brucei* is common in central Africa and is the causative agent of African sleeping sickness, a disease associated with severe chronic fatigue, coma, and can be fatal if left untreated.



Trypanosoma brucei, the causative agent of sleeping sickness, spends part of its life cycle in the tsetse fly and part in humans. (credit: modification of work by CDC)

Note:
Link to Learning

Watch this video to see *T. brucei* swimming.
https://www.openstaxcollege.org/l/T_brucei

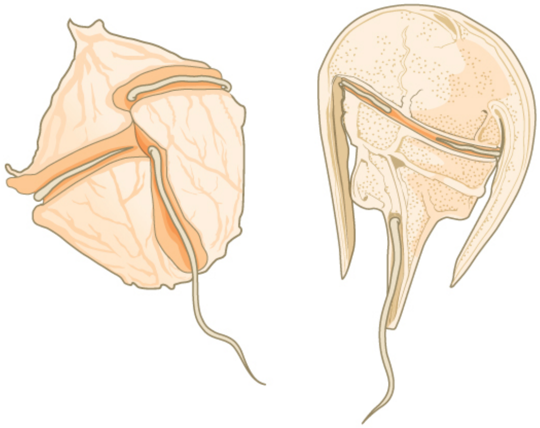
Chromalveolata

Current evidence suggests that species classified as chromalveolates are derived from a common ancestor that engulfed a photosynthetic red algal cell, which itself had already evolved chloroplasts from an endosymbiotic relationship with a photosynthetic prokaryote. Therefore, the ancestor of chromalveolates is believed to have resulted from a secondary endosymbiotic event. However, some chromalveolates appear to have lost red alga-derived plastid organelles or lack plastid genes altogether. Therefore, this supergroup should be considered a hypothesis-based working group that is subject to change. Chromalveolates include very important photosynthetic organisms, such as diatoms, brown algae, and significant disease agents in animals and plants. The chromalveolates can be subdivided into alveolates and stramenopiles.

Alveolates: Dinoflagellates, Apicomplexians, and Ciliates

A large body of data supports that the alveolates are derived from a shared common ancestor. The alveolates are named for the presence of an alveolus, or membrane-enclosed sac, beneath the cell membrane. The exact function of the alveolus is unknown, but it may be involved in osmoregulation. The alveolates are further categorized into some of the better-known protists: the dinoflagellates, the apicomplexans, and the ciliates.

Dinoflagellates exhibit extensive morphological diversity and can be photosynthetic, heterotrophic, or mixotrophic. Many dinoflagellates are encased in interlocking plates of cellulose. Two perpendicular flagella fit into the grooves between the cellulose plates, with one flagellum extending longitudinally and a second encircling the dinoflagellate ([link]). Together, the flagella contribute to the characteristic spinning motion of dinoflagellates. These protists exist in freshwater and marine habitats, and are a component of **plankton**, the typically microscopic organisms that drift through the water and serve as a crucial food source for larger aquatic organisms.

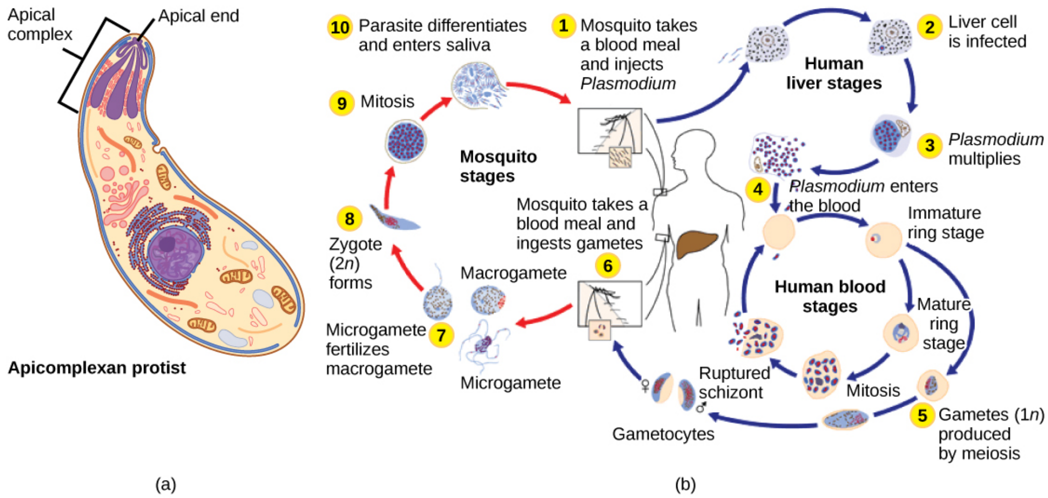


The dinoflagellates exhibit great diversity in shape. Many are encased in cellulose armor and have two flagella that fit in grooves between the plates. Movement of these two perpendicular flagella causes a spinning motion.

Some dinoflagellates generate light, called **bioluminescence**, when they are jarred or stressed. Large numbers of marine dinoflagellates (billions or trillions of cells per wave) can emit light and cause an entire breaking wave to twinkle or take on a brilliant blue color ([link]). For approximately 20 species of marine dinoflagellates, population explosions (also called blooms) during the summer months can tint the ocean with a muddy red color. This phenomenon is called a red tide, and it results from the abundant red pigments present in dinoflagellate plastids. In large quantities, these dinoflagellate species secrete an asphyxiating toxin that can kill fish, birds, and marine mammals. Red tides can be massively detrimental to commercial fisheries, and humans who consume these protists may become poisoned.

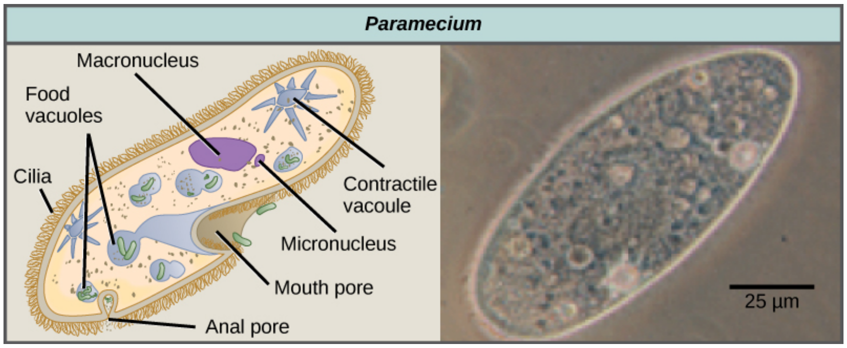
Bioluminescence is emitted from dinoflagellates in a breaking wave, as seen from the New Jersey coast. (credit: “catalano82”/Flickr)

The apicomplexan protists are so named because their microtubules, fibrin, and vacuoles are asymmetrically distributed at one end of the cell in a structure called an apical complex ([link]). The apical complex is specialized for entry and infection of host cells. Indeed, all apicomplexans are parasitic. This group includes the genus *Plasmodium*, which causes malaria in humans. Apicomplexan life cycles are complex, involving multiple hosts and stages of sexual and asexual reproduction.



(a) Apicomplexans are parasitic protists. They have a characteristic apical complex that enables them to infect host cells. (b) *Plasmodium*, the causative agent of malaria, has a complex life cycle typical of apicomplexans. (credit b: modification of work by CDC)

The ciliates, which include *Paramecium* and *Tetrahymena*, are a group of protists 10 to 3,000 micrometers in length that are covered in rows, tufts, or spirals of tiny cilia. By beating their cilia synchronously or in waves, ciliates can coordinate directed movements and ingest food particles. Certain ciliates have fused cilia-based structures that function like paddles, funnels, or fins. Ciliates also are surrounded by a pellicle, providing protection without compromising agility. The genus *Paramecium* includes protists that have organized their cilia into a plate-like primitive mouth, called an oral groove, which is used to capture and digest bacteria ([link]). Food captured in the oral groove enters a food vacuole, where it combines with digestive enzymes. Waste particles are expelled by an exocytic vesicle that fuses at a specific region on the cell membrane, called the anal pore. In addition to a vacuole-based digestive system, *Paramecium* also uses **contractile vacuoles**, which are osmoregulatory vesicles that fill with water as it enters the cell by osmosis and then contract to squeeze water from the cell.



Paramecium has a primitive mouth (called an oral groove) to ingest food, and an anal pore to excrete it. Contractile vacuoles allow the organism to excrete excess water. Cilia enable the organism to move. (credit "paramecium micrograph": modification of work by NIH; scale-bar data from Matt Russell)

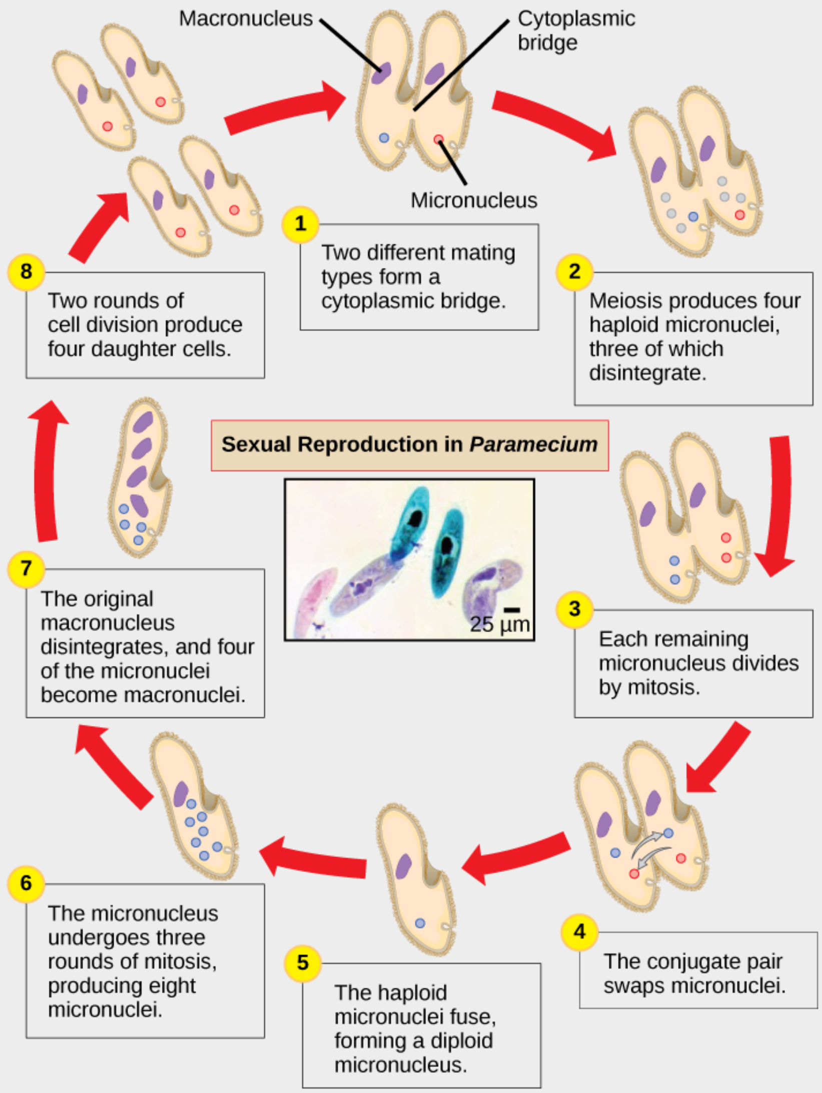
Note:
Link to Learning

Watch the video of the contractile vacuole of *Paramecium* expelling water to keep the cell osmotically balanced.
https://www.openstaxcollege.org/l/paramecium

Paramecium has two nuclei, a macronucleus and a micronucleus, in each cell. The micronucleus is essential for sexual reproduction, whereas the

macronucleus directs asexual binary fission and all other biological functions. The process of sexual reproduction in *Paramecium* underscores the importance of the micronucleus to these protists. *Paramecium* and most other ciliates reproduce sexually by conjugation. This process begins when two different mating types of *Paramecium* make physical contact and join with a cytoplasmic bridge ([link]). The diploid micronucleus in each cell then undergoes meiosis to produce four haploid micronuclei. Three of these degenerate in each cell, leaving one micronucleus that then undergoes mitosis, generating two haploid micronuclei. The cells each exchange one of these haploid nuclei and move away from each other. A similar process occurs in bacteria that have plasmids. Fusion of the haploid micronuclei generates a completely novel diploid pre-micronucleus in each conjugative cell. This pre-micronucleus undergoes three rounds of mitosis to produce eight copies, and the original macronucleus disintegrates. Four of the eight pre-micronuclei become full-fledged micronuclei, whereas the other four perform multiple rounds of DNA replication and go on to become new macronuclei. Two cell divisions then yield four new *Paramecia* from each original conjugative cell.

Note:
Art Connection



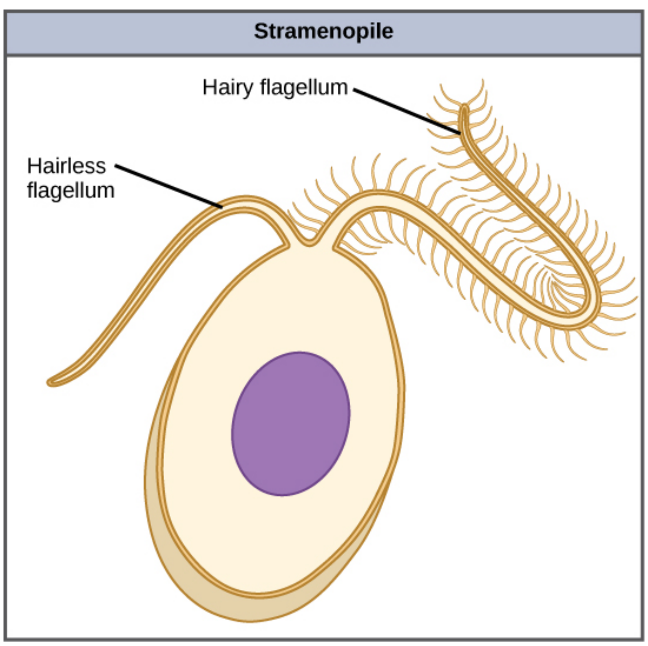
The complex process of sexual reproduction in *Paramecium* creates eight daughter cells from two original cells. Each cell has a macronucleus and a micronucleus. During sexual reproduction, the macronucleus dissolves and is replaced by a micronucleus. (credit “micrograph”: modification of work by Ian Sutton; scale-bar data from Matt Russell)

Which of the following statements about *Paramecium* sexual reproduction is false?

a. The macronuclei are derived from micronuclei.
b. Both mitosis and meiosis occur during sexual reproduction.
c. The conjugate pair swaps macronucleii.
d. Each parent produces four daughter cells.

Stramenopiles: Diatoms, Brown Algae, Golden Algae and Oomycetes

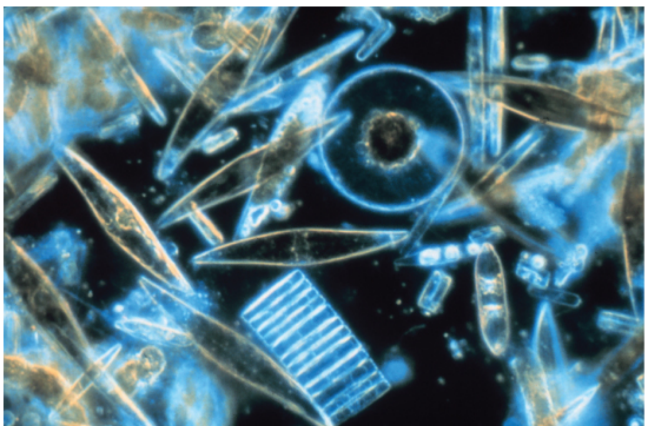
The other subgroup of chromalveolates, the stramenopiles, includes photosynthetic marine algae and heterotrophic protists. The unifying feature of this group is the presence of a textured, or “hairy,” flagellum. Many stramenopiles also have an additional flagellum that lacks hair-like projections ([link]). Members of this subgroup range in size from single-celled diatoms to the massive and multicellular kelp.



This stramenopile cell has a single

hairy flagellum and a secondary smooth flagellum.

The diatoms are unicellular photosynthetic protists that encase themselves in intricately patterned, glassy cell walls composed of silicon dioxide in a matrix of organic particles ([link]). These protists are a component of freshwater and marine plankton. Most species of diatoms reproduce asexually, although some instances of sexual reproduction and sporulation also exist. Some diatoms exhibit a slit in their silica shell, called a **raphe**. By expelling a stream of mucopolysaccharides from the raphe, the diatom can attach to surfaces or propel itself in one direction.



Assorted diatoms, visualized here using light microscopy, live among annual sea ice in McMurdo Sound, Antarctica. Diatoms range in size from 2 to 200 µm. (credit: Prof. Gordon T. Taylor, Stony Brook University, NSF, NOAA)

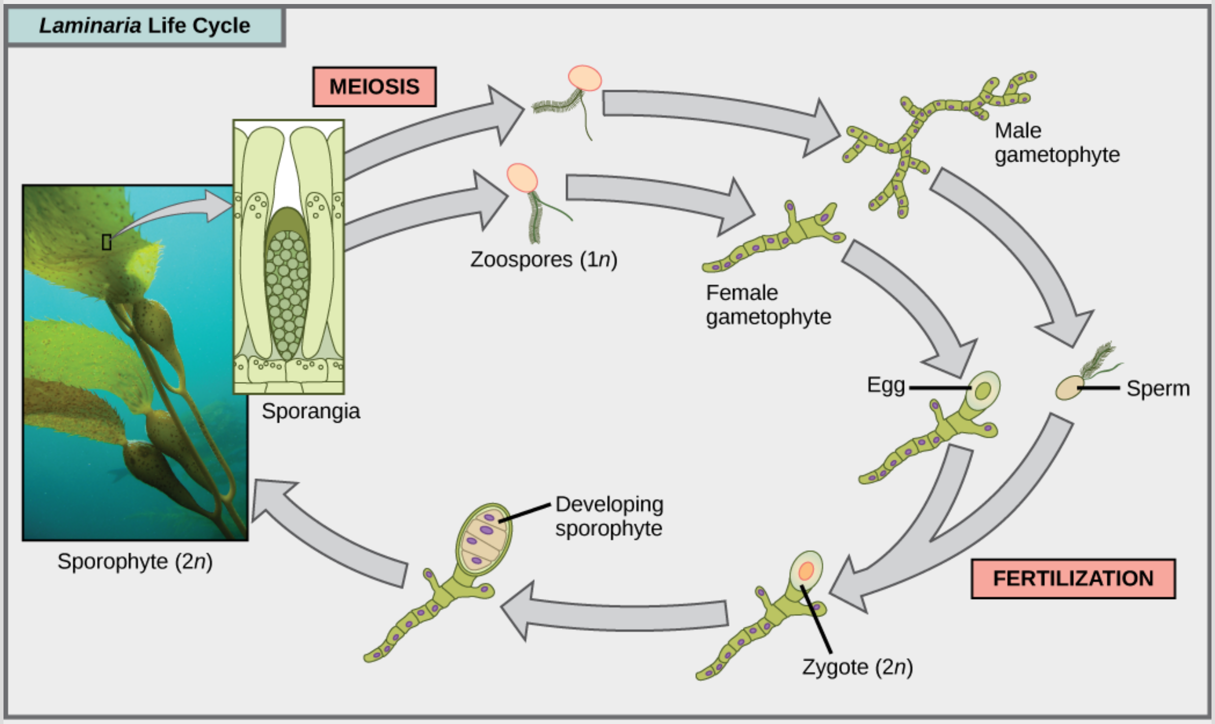
During periods of nutrient availability, diatom populations bloom to numbers greater than can be consumed by aquatic organisms. The excess diatoms die and sink to the sea floor where they are not easily reached by saprobes that feed on dead organisms. As a result, the carbon dioxide that the diatoms had consumed and incorporated into their cells during photosynthesis is not returned to the atmosphere. In general, this process by which carbon is transported deep into the ocean is described as the **biological carbon pump**, because carbon is “pumped” to the ocean depths where it is inaccessible to the atmosphere as carbon dioxide. The biological carbon pump is a crucial component of the carbon cycle that maintains lower atmospheric carbon dioxide levels.

Like diatoms, golden algae are largely unicellular, although some species can form large colonies. Their characteristic gold color results from their extensive use of carotenoids, a group of photosynthetic pigments that are generally yellow or orange in color. Golden algae are found in both freshwater and marine environments, where they form a major part of the plankton community.

The brown algae are primarily marine, multicellular organisms that are known colloquially as seaweeds. Giant kelps are a type of brown algae. Some brown algae have evolved specialized tissues that resemble terrestrial plants, with root-like holdfasts, stem-like stipes, and leaf-like blades that are capable of photosynthesis. The stipes of giant kelps are enormous, extending in some cases for 60 meters. A variety of algal life cycles exists, but the most complex is alternation of generations, in which both haploid and diploid stages involve multicellularity. Compare this life cycle to that of humans, for instance. Haploid gametes produced by meiosis (sperm and egg) combine in fertilization to generate a diploid zygote that undergoes many rounds of mitosis to produce a multicellular embryo and then a fetus. However, the individual sperm and egg themselves never become multicellular beings. Terrestrial plants also have evolved alternation of generations. In the brown algae genus *Laminaria*, haploid spores develop into multicellular gametophytes, which produce haploid gametes that combine to produce diploid organisms that then become multicellular organisms with a different structure from the haploid form ([link]). Certain

other organisms perform alternation of generations in which both the haploid and diploid forms look the same.

Note:
Art Connection



Several species of brown algae, such as the *Laminaria* shown here, have evolved life cycles in which both the haploid (gametophyte) and diploid (sporophyte) forms are multicellular. The gametophyte is different in structure than the sporophyte. (credit “laminaria photograph”: modification of work by Claire Fackler, CINMS, NOAA Photo Library)

Which of the following statements about the *Laminaria* life cycle is false?

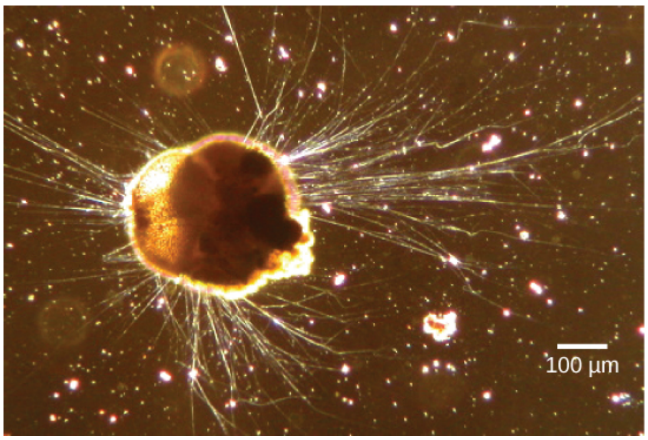
a. 1*n* zoospores form in the sporangia.
b. The sporophyte is the 2*n* plant.
c. The gametophyte is diploid.
d. Both the gametophyte and sporophyte stages are multicellular.

The water molds, oomycetes (“egg fungus”), were so-named based on their fungus-like morphology, but molecular data have shown that the water molds are not closely related to fungi. The oomycetes are characterized by a cellulose-based cell wall and an extensive network of filaments that allow for nutrient uptake. As diploid spores, many oomycetes have two oppositely directed flagella (one hairy and one smooth) for locomotion. The oomycetes are nonphotosynthetic and include many saprobes and parasites. The saprobes appear as white fluffy growths on dead organisms ([link]). Most oomycetes are aquatic, but some parasitize terrestrial plants. One plant pathogen is *Phytophthora infestans*, the causative agent of late blight of potatoes, such as occurred in the nineteenth century Irish potato famine.

A saprobic oomycete engulfs a dead insect. (credit: modification of work by Thomas Bresson)

Rhizaria

The Rhizaria supergroup includes many of the amoebas, most of which have threadlike or needle-like pseudopodia ([link]). Pseudopodia function to trap and engulf food particles and to direct movement in rhizarian protists. These pseudopods project outward from anywhere on the cell surface and can anchor to a substrate. The protist then transports its cytoplasm into the pseudopod, thereby moving the entire cell. This type of motion, called **cytoplasmic streaming**, is used by several diverse groups of protists as a means of locomotion or as a method to distribute nutrients and oxygen.



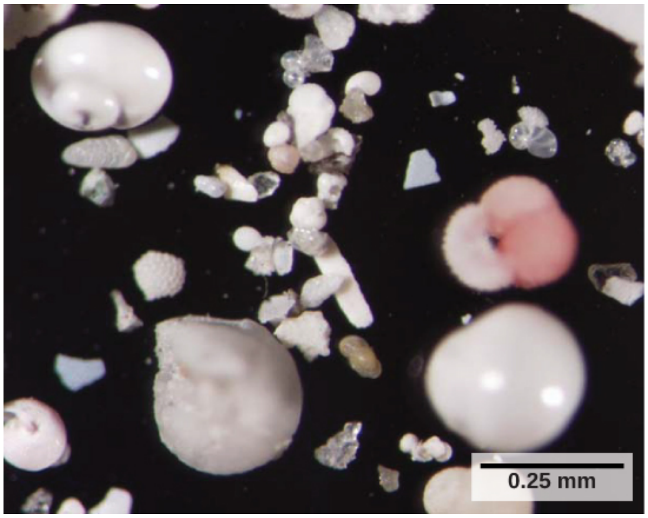
Ammonia tepida, a Rhizaria species viewed here using phase contrast light microscopy, exhibits many threadlike pseudopodia. (credit: modification of work by Scott Fay, UC Berkeley; scale-bar data from Matt Russell)

Note:
Link to Learning

Take a look at this video to see cytoplasmic streaming in a green alga. https://www.openstaxcollege.org/l/chara_corallina

Forams

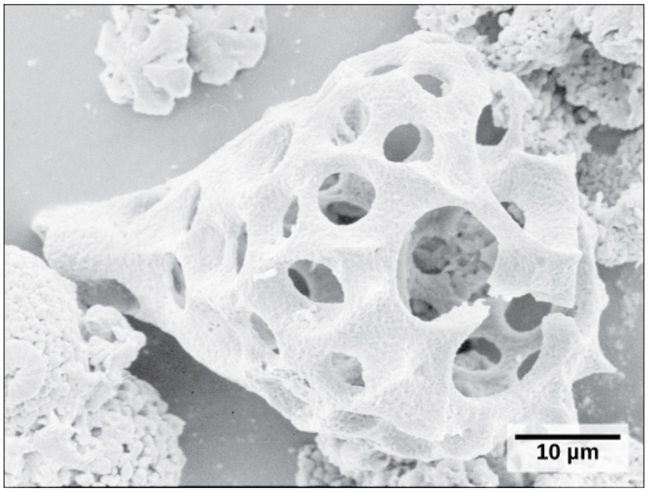
Foraminiferans, or forams, are unicellular heterotrophic protists, ranging from approximately 20 micrometers to several centimeters in length, and occasionally resembling tiny snails ([link]). As a group, the forams exhibit porous shells, called **tests** that are built from various organic materials and typically hardened with calcium carbonate. The tests may house photosynthetic algae, which the forams can harvest for nutrition. Foram pseudopodia extend through the pores and allow the forams to move, feed, and gather additional building materials. Typically, forams are associated with sand or other particles in marine or freshwater habitats. Foraminiferans are also useful as indicators of pollution and changes in global weather patterns.



These shells from foraminifera sank to the sea floor. (credit: Deep East 2001, NOAA/OER)

Radiolarians

A second subtype of Rhizaria, the radiolarians, exhibit intricate exteriors of glassy silica with radial or bilateral symmetry ([link]). Needle-like pseudopods supported by microtubules radiate outward from the cell bodies of these protists and function to catch food particles. The shells of dead radiolarians sink to the ocean floor, where they may accumulate in 100 meter-thick depths. Preserved, sedimented radiolarians are very common in the fossil record.



This fossilized radiolarian shell was imaged using a scanning electron microscope. (credit: modification of work by Hannes Grobe, Alfred Wegener Institute; scale-bar data from Matt Russell)

Archaeplastida

Red algae and green algae are included in the supergroup Archaeplastida. It was from a common ancestor of these protists that the land plants evolved, since their closest relatives are found in this group. Molecular evidence supports that all Archaeplastida are descendents of an endosymbiotic relationship between a heterotrophic protist and a cyanobacterium. The red and green algae include unicellular, multicellular, and colonial forms.

Red Algae

Red algae, or rhodophytes, are primarily multicellular, lack flagella, and range in size from microscopic, unicellular protists to large, multicellular forms grouped into the informal seaweed category. The red algae life cycle

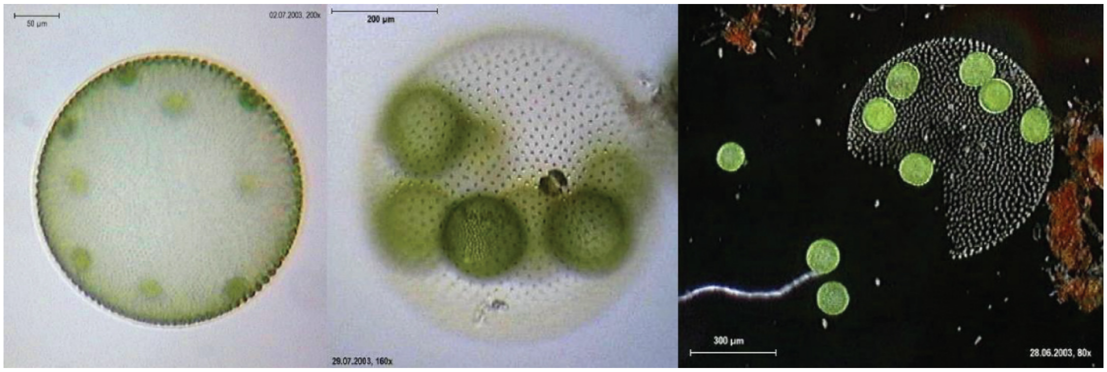
is an alternation of generations. Some species of red algae contain phycoerythrins, photosynthetic accessory pigments that are red in color and outcompete the green tint of chlorophyll, making these species appear as varying shades of red. Other protists classified as red algae lack phycoerythrins and are parasites. Red algae are common in tropical waters where they have been detected at depths of 260 meters. Other red algae exist in terrestrial or freshwater environments.

Green Algae: Chlorophytes and Charophytes

The most abundant group of algae is the green algae. The green algae exhibit similar features to the land plants, particularly in terms of chloroplast structure. That this group of protists shared a relatively recent common ancestor with land plants is well supported. The green algae are subdivided into the chlorophytes and the charophytes. The charophytes are the closest living relatives to land plants and resemble them in morphology and reproductive strategies. Charophytes are common in wet habitats, and their presence often signals a healthy ecosystem.

The chlorophytes exhibit great diversity of form and function. Chlorophytes primarily inhabit freshwater and damp soil, and are a common component of plankton. *Chlamydomonas* is a simple, unicellular chlorophyte with a pear-shaped morphology and two opposing, anterior flagella that guide this protist toward light sensed by its eyespot. More complex chlorophyte species exhibit haploid gametes and spores that resemble *Chlamydomonas*.

The chlorophyte *Volvox* is one of only a few examples of a colonial organism, which behaves in some ways like a collection of individual cells, but in other ways like the specialized cells of a multicellular organism ([link]). *Volvox* colonies contain 500 to 60,000 cells, each with two flagella, contained within a hollow, spherical matrix composed of a gelatinous glycoprotein secretion. Individual *Volvox* cells move in a coordinated fashion and are interconnected by cytoplasmic bridges. Only a few of the cells reproduce to create daughter colonies, an example of basic cell specialization in this organism.



Volvox aureus is a green alga in the supergroup Archaeplastida. This species exists as a colony, consisting of cells immersed in a gel-like matrix and intertwined with each other via hair-like cytoplasmic extensions. (credit: Dr. Ralf Wagner)

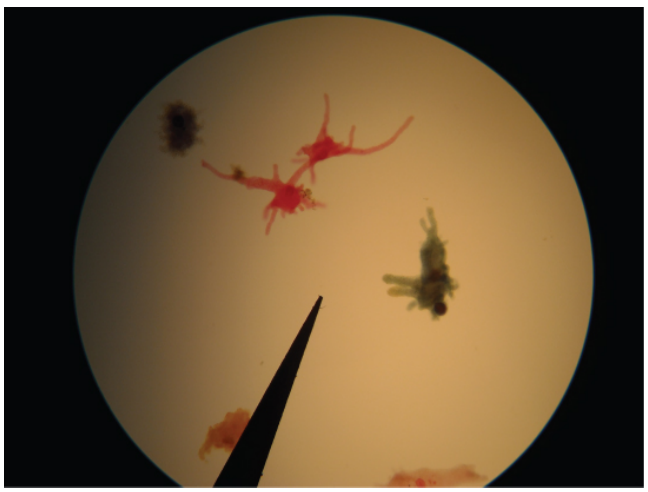
True multicellular organisms, such as the sea lettuce, *Ulva*, are represented among the chlorophytes. In addition, some chlorophytes exist as large, multinucleate, single cells. Species in the genus *Caulerpa* exhibit flattened fern-like foliage and can reach lengths of 3 meters ([link]). *Caulerpa* species undergo nuclear division, but their cells do not complete cytokinesis, remaining instead as massive and elaborate single cells.

Caulerpa taxifolia is a chlorophyte consisting of a single cell

containing potentially thousands of nuclei. (credit: NOAA)

Amoebozoa

The amoebozoans characteristically exhibit pseudopodia that extend like tubes or flat lobes, rather than the hair-like pseudopodia of rhizarian amoeba ([link]). The Amoebozoa include several groups of unicellular amoeba-like organisms that are free-living or parasites.

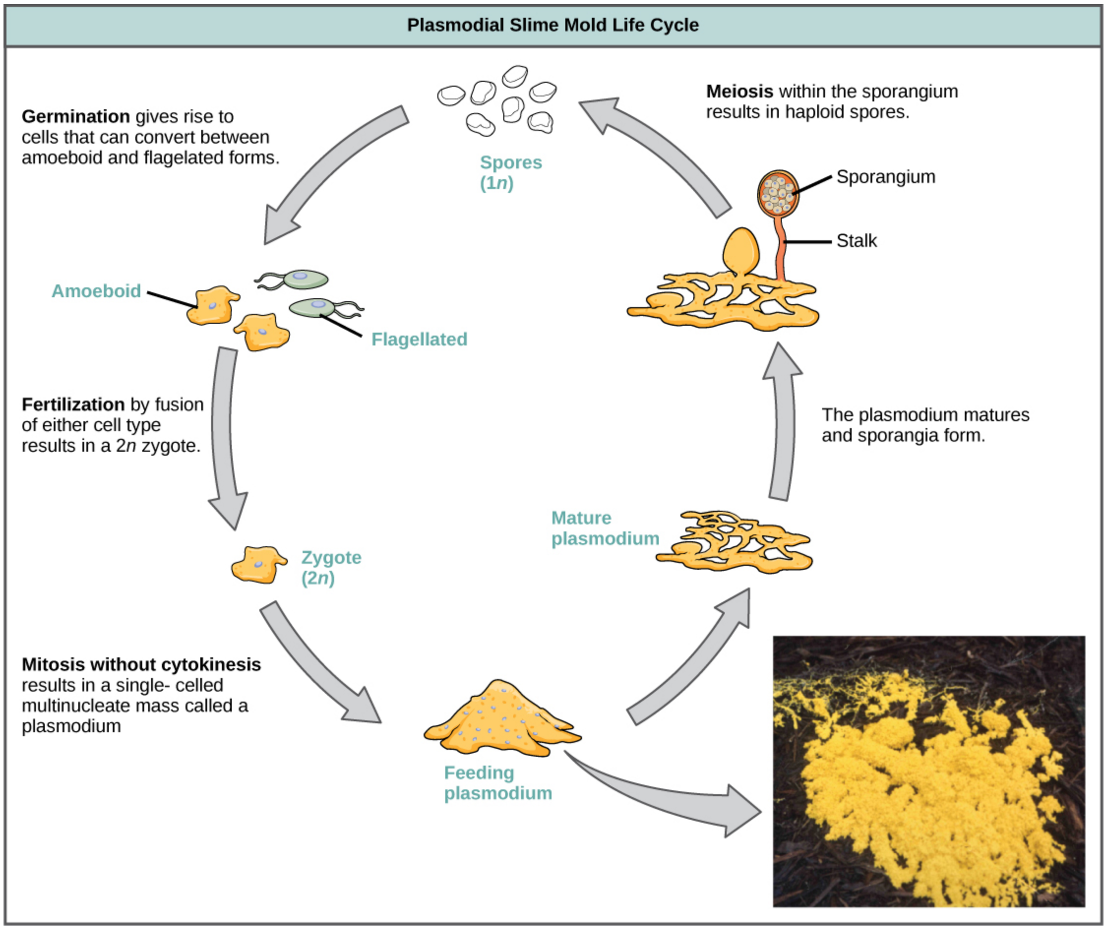


Amoebae with tubular and lobe-shaped pseudopodia are seen under a microscope. These isolates would be morphologically classified as amoebozoans.

Slime Molds

A subset of the amoebozoans, the slime molds, has several morphological similarities to fungi that are thought to be the result of convergent evolution. For instance, during times of stress, some slime molds develop into spore-generating fruiting bodies, much like fungi.

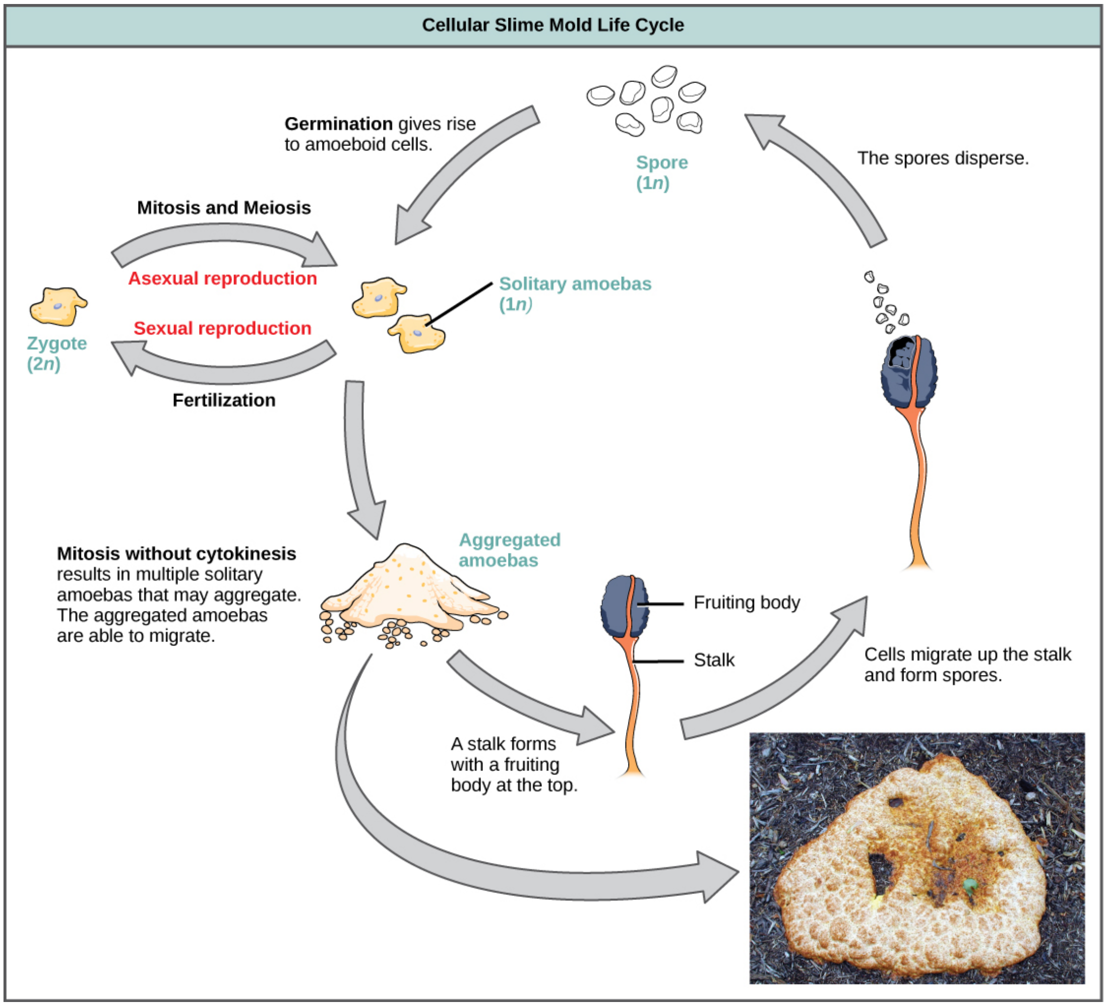
The slime molds are categorized on the basis of their life cycles into plasmodial or cellular types. Plasmodial slime molds are composed of large, multinucleate cells and move along surfaces like an amorphous blob of slime during their feeding stage ([link]). Food particles are lifted and engulfed into the slime mold as it glides along. Upon maturation, the plasmodium takes on a net-like appearance with the ability to form fruiting bodies, or sporangia, during times of stress. Haploid spores are produced by meiosis within the sporangia, and spores can be disseminated through the air or water to potentially land in more favorable environments. If this occurs, the spores germinate to form ameboid or flagellate haploid cells that can combine with each other and produce a diploid zygotic slime mold to complete the life cycle.



The life cycle of the plasmodial slime mold is shown. The brightly colored plasmodium in the inset photo is a single-celled, multinucleate mass. (credit: modification of work by Dr. Jonatha Gott and the Center for RNA Molecular Biology, Case Western Reserve University)

The cellular slime molds function as independent amoeboid cells when nutrients are abundant ([link]). When food is depleted, cellular slime molds pile onto each other into a mass of cells that behaves as a single unit, called a slug. Some cells in the slug contribute to a 2–3-millimeter stalk, drying up and dying in the process. Cells atop the stalk form an asexual fruiting body that contains haploid spores. As with plasmodial slime molds, the spores are disseminated and can germinate if they land in a moist environment. One

representative genus of the cellular slime molds is *Dictyostelium*, which commonly exists in the damp soil of forests.



Cellular slime molds may exist as solitary or aggregated amoebas. (credit: modification of work by “thatredhead4”/Flickr)

Note:
Link to Learning

View this video to see the formation of a fruiting body by a cellular slime mold.
https://www.openstaxcollege.org/l/slime_mold

Opisthokonta

The opisthokonts include the animal-like choanoflagellates, which are believed to resemble the common ancestor of sponges and, in fact, all animals. Choanoflagellates include unicellular and colonial forms, and number about 244 described species. These organisms exhibit a single, apical flagellum that is surrounded by a contractile collar composed of microvilli. The collar uses a similar mechanism to sponges to filter out bacteria for ingestion by the protist. The morphology of choanoflagellates was recognized early on as resembling the collar cells of sponges, and suggesting a possible relationship to animals.

The Mesomycetozoa form a small group of parasites, primarily of fish, and at least one form that can parasitize humans. Their life cycles are poorly understood. These organisms are of special interest, because they appear to be so closely related to animals. In the past, they were grouped with fungi and other protists based on their morphology.

Section Summary

The process of classifying protists into meaningful groups is ongoing, but genetic data in the past 20 years have clarified many relationships that were previously unclear or mistaken. The majority view at present is to order all eukaryotes into six supergroups: Excavata, Chromalveolata, Rhizaria, Archaeplastida, Amoebozoa, and Opisthokonta. The goal of this

classification scheme is to create clusters of species that all are derived from a common ancestor. At present, the monophyly of some of the supergroups are better supported by genetic data than others. Although tremendous variation exists within the supergroups, commonalities at the morphological, physiological, and ecological levels can be identified.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about *Paramecium* sexual reproduction is false?

a. The macronuclei are derived from micronuclei.
b. Both mitosis and meiosis occur during sexual reproduction.
c. The conjugate pair swaps macronuclei.
d. Each parent produces four daughter cells.

Solution:

[link] C

Exercise:

Problem:

[link] Which of the following statements about the *Laminaria* life cycle is false?

a. 1*n* zoospores form in the sporangia.
b. The sporophyte is the *2n* plant.
c. The gametophyte is diploid.
d. Both the gametophyte and sporophyte stages are multicellular.

Solution:

[link] C

Review Questions

Exercise:

Problem:

Which protist group exhibits mitochondrial remnants with reduced functionality?

a. slime molds
b. diatoms
c. parabasalids
d. dinoflagellates

Solution:

C

Exercise:

Problem:

Conjugation between two *Paramecia* produces ________ total daughter cells.

a. 2
b. 4
c. 8
d. 16

Solution:

C

Exercise:

Problem: What is the function of the raphe in diatoms?

a. locomotion
b. defense
c. capturing food
d. photosynthesis

Solution:

A

Exercise:

Problem:

What genus of protists appears to contradict the statement that unicellularity restricts cell size?

a. *Dictyostelium*
b. *Ulva*
c. *Plasmodium*
d. *Caulerpa*

Solution:

D

Free Response

Exercise:

Problem:

The chlorophyte (green algae) genera *Ulva* and *Caulerpa* both have macroscopic leaf-like and stem-like structures, but only *Ulva* species are considered truly multicellular. Explain why.

Solution:

Unlike *Ulva*, protists in the genus *Caulerpa* actually are large, multinucleate, single cells. Because these organisms undergo mitosis without cytokinesis and lack cytoplasmic divisions, they cannot be considered truly multicellular.

Exercise:

Problem:

Why might a light-sensing eyespot be ineffective for an obligate saprobe? Suggest an alternative organ for a saprobic protist.

Solution:

By definition, an obligate saprobe lacks the ability to perform photosynthesis, so it cannot directly obtain nutrition by searching for light. Instead, a chemotactic mechanism that senses the odors released during decay might be a more effective sensing organ for a saprobe.

Glossary

biological carbon pump
: process by which inorganic carbon is fixed by photosynthetic species that then die and fall to the sea floor where they cannot be reached by saprobes and their carbon dioxide consumption cannot be returned to the atmosphere

bioluminescence
: generation and emission of light by an organism, as in dinoflagellates

contractile vacuole
: vesicle that fills with water (as it enters the cell by osmosis) and then contracts to squeeze water from the cell; an osmoregulatory vesicle

cytoplasmic streaming

movement of cytoplasm into an extended pseudopod such that the entire cell is transported to the site of the pseudopod

hydrogenosome
: organelle carried by parabasalids (Excavata) that functions anaerobically and outputs hydrogen gas as a byproduct; likely evolved from mitochondria

kinetoplast
: mass of DNA carried within the single, oversized mitochondrion, characteristic of kinetoplastids (phylum: Euglenozoa)

mitosome
: nonfunctional organelle carried in the cells of diplomonads (Excavata) that likely evolved from a mitochondrion

plankton
: diverse group of mostly microscopic organisms that drift in marine and freshwater systems and serve as a food source for larger aquatic organisms

raphe
: slit in the silica shell of diatoms through which the protist secretes a stream of mucopolysaccharides for locomotion and attachment to substrates

test
: porous shell of a foram that is built from various organic materials and typically hardened with calcium carbonate

Ecology of Protists
By the end of this section, you will be able to:

- Describe the role that protists play in the ecosystem
- Describe important pathogenic species of protists

Protists function in various ecological niches. Whereas some protist species are essential components of the food chain and generators of biomass, others function in the decomposition of organic materials. Still other protists are dangerous human pathogens or causative agents of devastating plant diseases.

Primary Producers/Food Sources

Protists are essential sources of nutrition for many other organisms. In some cases, as in plankton, protists are consumed directly. Alternatively, photosynthetic protists serve as producers of nutrition for other organisms. For instance, photosynthetic dinoflagellates called zooxanthellae use sunlight to fix inorganic carbon. In this symbiotic relationship, these protists provide nutrients for coral polyps ([link]) that house them, giving corals a boost of energy to secrete a calcium carbonate skeleton. In turn, the corals provide the protist with a protected environment and the compounds needed for photosynthesis. This type of symbiotic relationship is important in nutrient-poor environments. Without dinoflagellate symbionts, corals lose algal pigments in a process called coral bleaching, and they eventually die. This explains why reef-building corals do not reside in waters deeper than 20 meters: insufficient light reaches those depths for dinoflagellates to photosynthesize.

Coral polyps obtain nutrition through a symbiotic relationship with dinoflagellates.

The protists themselves and their products of photosynthesis are essential—directly or indirectly—to the survival of organisms ranging from bacteria to mammals ([link]). As primary producers, protists feed a large proportion of the world's aquatic species. (On land, terrestrial plants serve as primary producers.) In fact, approximately one-quarter of the world's photosynthesis is conducted by protists, particularly dinoflagellates, diatoms, and multicellular algae.

Virtually all aquatic organisms depend directly or indirectly on protists for food. (credit “mollusks”: modification of work by Craig Stihler, USFWS; credit “crab”: modification of work by David Berkowitz; credit “dolphin”: modification of work by Mike Baird; credit “fish”: modification of work by Tim Sheerman-Chase; credit “penguin”: modification of work by Aaron Logan)

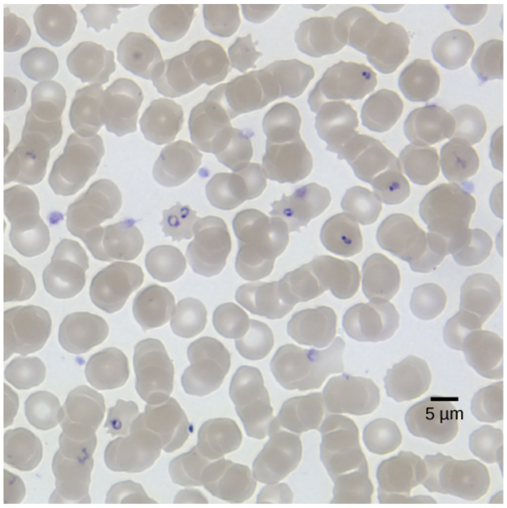
Protists do not create food sources only for sea-dwelling organisms. For instance, certain anaerobic parabasalid species exist in the digestive tracts of termites and wood-eating cockroaches, where they contribute an essential step in the digestion of cellulose ingested by these insects as they bore through wood.

Human Pathogens

A pathogen is anything that causes disease. Parasites live in or on an organism and harm the organism. A significant number of protists are pathogenic parasites that must infect other organisms to survive and propagate. Protist parasites include the causative agents of malaria, African sleeping sickness, and waterborne gastroenteritis in humans. Other protist pathogens prey on plants, effecting massive destruction of food crops.

Plasmodium Species

Members of the genus *Plasmodium* must colonize both a mosquito and a vertebrate to complete their life cycle. In vertebrates, the parasite develops in liver cells and goes on to infect red blood cells, bursting from and destroying the blood cells with each asexual replication cycle ([link]). Of the four *Plasmodium* species known to infect humans, *P. falciparum* accounts for 50 percent of all malaria cases and is the primary cause of disease-related fatalities in tropical regions of the world. In 2010, it was estimated that malaria caused between one-half and one million deaths, mostly in African children. During the course of malaria, *P. falciparum* can infect and destroy more than one-half of a human’s circulating blood cells, leading to severe anemia. In response to waste products released as the parasites burst from infected blood cells, the host immune system mounts a massive inflammatory response with episodes of delirium-inducing fever as parasites lyse red blood cells, spilling parasite waste into the bloodstream. *P. falciparum* is transmitted to humans by the African malaria mosquito, *Anopheles gambiae*. Techniques to kill, sterilize, or avoid exposure to this highly aggressive mosquito species are crucial to malaria control.



Red blood cells are shown to be infected with *P. falciparum*, the causative agent of malaria. In this light microscopic image taken using a 100× oil immersion lens, the ring-shaped *P. falciparum* stains purple. (credit: modification of work by Michael Zahniser; scale-bar data from Matt Russell)

Note:
Link to Learning

This movie depicts the pathogenesis of *Plasmodium falciparum*, the causative agent of malaria.

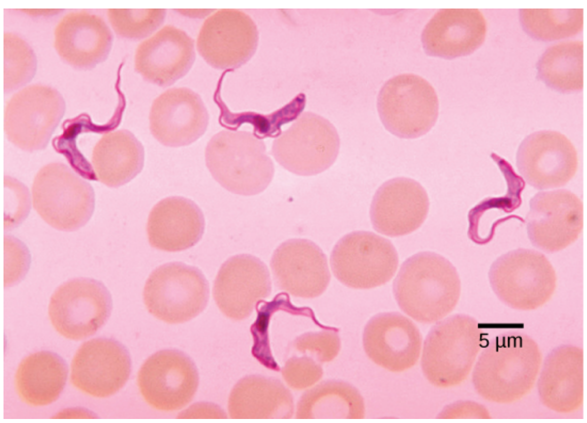
Trypanosomes

Trypanosoma brucei, the parasite that is responsible for African sleeping sickness, confounds the human immune system by changing its thick layer of surface glycoproteins with each infectious cycle ([link]). The glycoproteins are identified by the immune system as foreign antigens, and a specific antibody defense is mounted against the parasite. However, *T. brucei* has thousands of possible antigens, and with each subsequent generation, the protist switches to a glycoprotein coating with a different molecular structure. In this way, *T. brucei* is capable of replicating continuously without the immune system ever succeeding in clearing the parasite. Without treatment, *T. brucei* crosses the blood-brain barrier and infects the central nervous system, causing the patient to lapse into a coma and eventually die. During epidemic periods, mortality from the disease can be high. Greater surveillance and control measures lead to a reduction in reported cases; some of the lowest numbers reported in 50 years (fewer than 10,000 cases in all of sub-Saharan Africa) have happened since 2009.

Note:
Link to Learning

This movie discusses the pathogenesis of *Trypanosoma brucei*, the causative agent of African sleeping sickness.

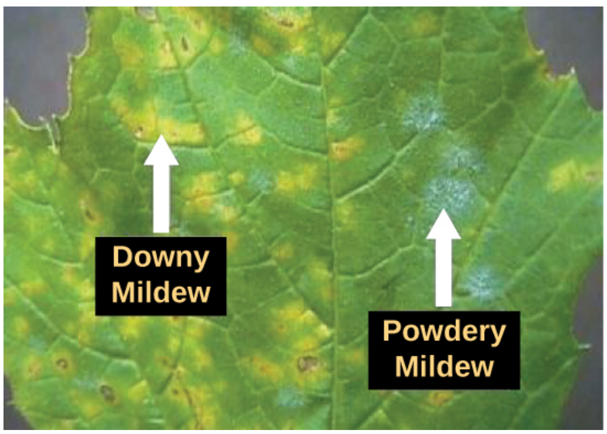
In Latin America, another species, *T. cruzi*, is responsible for Chagas disease. *T. cruzi* infections are mainly caused by a blood-sucking bug. The parasite inhabits heart and digestive system tissues in the chronic phase of infection, leading to malnutrition and heart failure due to abnormal heart rhythms. An estimated 10 million people are infected with Chagas disease, and it caused 10,000 deaths in 2008.



Trypanosomes are shown among red blood cells. (credit: modification of work by Dr. Myron G. Shultz; scale-bar data from Matt Russell)

Plant Parasites

Protist parasites of terrestrial plants include agents that destroy food crops. The oomycete *Plasmopara viticola* parasitizes grape plants, causing a disease called downy mildew ([link]). Grape plants infected with *P. viticola* appear stunted and have discolored, withered leaves. The spread of downy mildew nearly collapsed the French wine industry in the nineteenth century.



Both downy and powdery mildews on this grape leaf are caused by an infection of *P. viticola*. (credit: modification of work by USDA)

Phytophthora infestans is an oomycete responsible for potato late blight, which causes potato stalks and stems to decay into black slime ([link]). Widespread potato blight caused by *P. infestans* precipitated the well-known Irish potato famine in the nineteenth century that claimed the lives of approximately 1 million people and led to the emigration of at least 1 million more from Ireland. Late blight continues to plague potato crops in certain parts of the United States and Russia, wiping out as much as 70 percent of crops when no pesticides are applied.

These unappetizing remnants result from an infection with *P. infestans*, the causative agent of potato late blight. (credit: USDA)

Agents of Decomposition

The fungus-like protist saprobes are specialized to absorb nutrients from nonliving organic matter, such as dead organisms or their wastes. For instance, many types of oomycetes grow on dead animals or algae. Saprobic protists have the essential function of returning inorganic nutrients to the soil and water. This process allows for new plant growth, which in turn generates sustenance for other organisms along the food chain. Indeed, without saprobe species, such as protists, fungi, and bacteria, life would cease to exist as all organic carbon became "tied up" in dead organisms.

Section Summary

Protists function at several levels of the ecological food web: as primary producers, as direct food sources, and as decomposers. In addition, many protists are parasites of plants and animals that can cause deadly human diseases or destroy valuable crops.

Review Questions

Exercise:

Problem: An example of carbon fixation is ______________.

a. photosynthesis
b. decomposition
c. phagocytosis
d. parasitism

Solution:

A

Exercise:

Problem:

Which parasitic protist evades the host immune system by altering its surface proteins with each generation?

a. *Paramecium caudatum*
b. *Trypanosoma brucei*
c. *Plasmodium falciparum*
d. *Phytophthora infestans*

Solution:

B

Free Response

Exercise:

Problem:

How does killing *Anopheles* mosquitoes affect the *Plasmodium* protists?

Solution:

Plasmodium parasites infect humans and cause malaria. However, they must complete part of their life cycle within *Anopheles* mosquitoes, and they can only infect humans via the bite wound of a mosquito. If the mosquito population is decreased, then fewer *Plasmodium* would be able to develop and infect humans, thereby reducing the incidence of human infections with this parasite.

Exercise:

Problem:

Without treatment, why does African sleeping sickness invariably lead to death?

Solution:

The trypanosomes that cause this disease are capable of expressing a glycoprotein coat with a different molecular structure with each generation. Because the immune system must respond to specific antigens to raise a meaningful defense, the changing nature of trypanosome antigens prevents the immune system from ever clearing this infection. Massive trypanosome infection eventually leads to host organ failure and death.

Characteristics of Fungi

By the end of this section, you will be able to:

- List the characteristics of fungi
- Describe the composition of the mycelium
- Describe the mode of nutrition of fungi
- Explain sexual and asexual reproduction in fungi

Although humans have used yeasts and mushrooms since prehistoric times, until recently, the biology of fungi was poorly understood. Up until the mid-20th century, many scientists classified fungi as plants. Fungi, like plants, arose mostly sessile and seemingly rooted in place. They possess a stem-like structure similar to plants, as well as having a root-like fungal mycelium in the soil. In addition, their mode of nutrition was poorly understood. Progress in the field of fungal biology was the result of **mycology**: the scientific study of fungi. Based on fossil evidence, fungi appeared in the pre-Cambrian era, about 450 million years ago. Molecular biology analysis of the fungal genome demonstrates that fungi are more closely related to animals than plants. They are a polyphyletic group of organisms that share characteristics, rather than sharing a single common ancestor.

Note:
Career Connection
Mycologist
Mycologists are biologists who study fungi. Mycology is a branch of microbiology, and many mycologists start their careers with a degree in microbiology. To become a mycologist, a bachelor's degree in a biological science (preferably majoring in microbiology) and a master's degree in mycology are minimally necessary. Mycologists can specialize in taxonomy and fungal genomics, molecular and cellular biology, plant pathology, biotechnology, or biochemistry. Some medical microbiologists concentrate on the study of infectious diseases caused by fungi (mycoses). Mycologists collaborate with zoologists and plant pathologists to identify and control difficult fungal infections, such as the devastating chestnut blight, the mysterious decline in frog populations in many areas of the

world, or the deadly epidemic called white nose syndrome, which is decimating bats in the Eastern United States.
Government agencies hire mycologists as research scientists and technicians to monitor the health of crops, national parks, and national forests. Mycologists are also employed in the private sector by companies that develop chemical and biological control products or new agricultural products, and by companies that provide disease control services. Because of the key role played by fungi in the fermentation of alcohol and the preparation of many important foods, scientists with a good understanding of fungal physiology routinely work in the food technology industry. Oenology, the science of wine making, relies not only on the knowledge of grape varietals and soil composition, but also on a solid understanding of the characteristics of the wild yeasts that thrive in different wine-making regions. It is possible to purchase yeast strains isolated from specific grape-growing regions. The great French chemist and microbiologist, Louis Pasteur, made many of his essential discoveries working on the humble brewer's yeast, thus discovering the process of fermentation.

Cell Structure and Function

Fungi are eukaryotes, and as such, have a complex cellular organization. As eukaryotes, fungal cells contain a membrane-bound nucleus. The DNA in the nucleus is wrapped around histone proteins, as is observed in other eukaryotic cells. A few types of fungi have structures comparable to bacterial plasmids (loops of DNA); however, the horizontal transfer of genetic information from one mature bacterium to another rarely occurs in fungi. Fungal cells also contain mitochondria and a complex system of internal membranes, including the endoplasmic reticulum and Golgi apparatus.

Unlike plant cells, fungal cells do not have chloroplasts or chlorophyll. Many fungi display bright colors arising from other cellular pigments, ranging from red to green to black. The poisonous *Amanita muscaria* (fly agaric) is recognizable by its bright red cap with white patches ([link]).

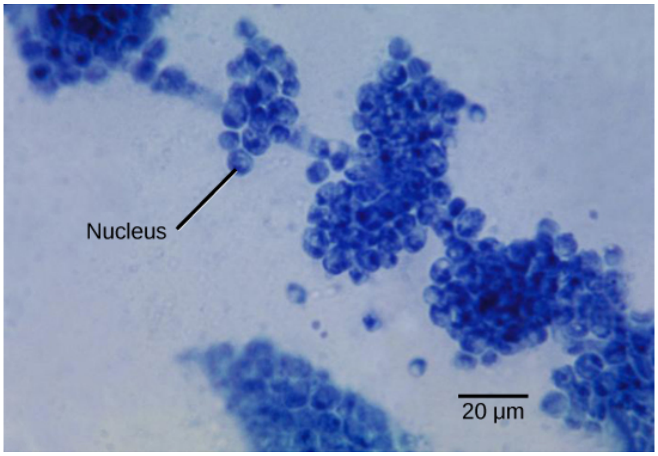
Pigments in fungi are associated with the cell wall and play a protective role against ultraviolet radiation. Some fungal pigments are toxic.

The poisonous *Amanita muscaria* is native to temperate and boreal regions of North America. (credit: Christine Majul)

Like plant cells, fungal cells have a thick cell wall. The rigid layers of fungal cell walls contain complex polysaccharides called chitin and glucans. Chitin, also found in the exoskeleton of insects, gives structural strength to the cell walls of fungi. The wall protects the cell from desiccation and predators. Fungi have plasma membranes similar to other eukaryotes, except that the structure is stabilized by ergosterol: a steroid molecule that replaces the cholesterol found in animal cell membranes. Most members of the kingdom Fungi are nonmotile. Flagella are produced only by the gametes in the primitive Phylum Chytridiomycota.

Growth

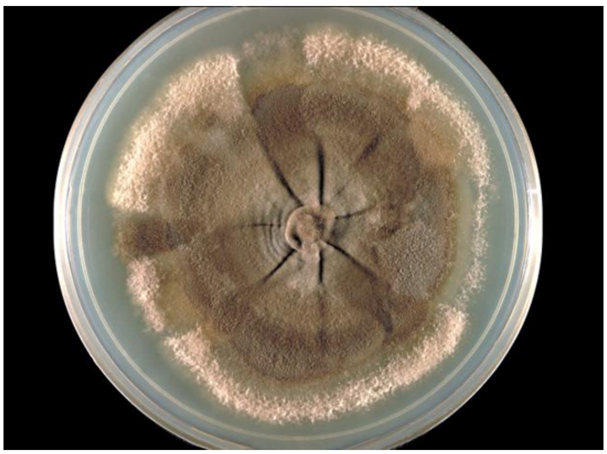
The vegetative body of a fungus is a unicellular or multicellular **thallus**. Dimorphic fungi can change from the unicellular to multicellular state depending on environmental conditions. Unicellular fungi are generally referred to as **yeasts**. *Saccharomyces cerevisiae* (baker's yeast) and *Candida* species (the agents of thrush, a common fungal infection) are examples of unicellular fungi ([link]).



Candida albicans is a yeast cell and the agent of candidiasis and thrush. This organism has a similar morphology to coccus bacteria; however, yeast is a eukaryotic organism (note the nucleus). (credit: modification of work by Dr. Godon Roberstad, CDC; scale-bar data from Matt Russell)

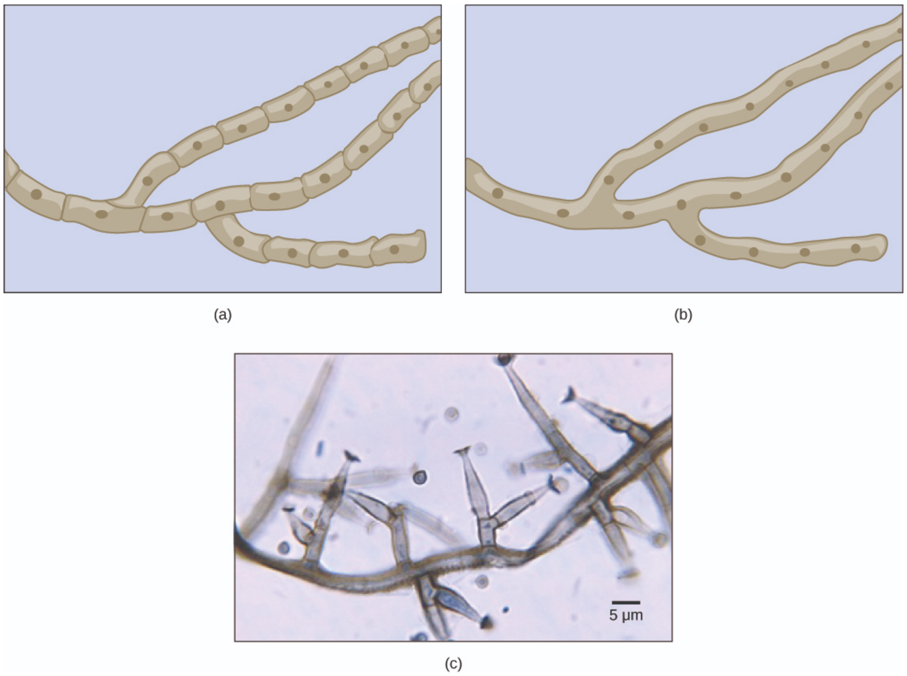
Most fungi are multicellular organisms. They display two distinct morphological stages: the vegetative and reproductive. The vegetative stage consists of a tangle of slender thread-like structures called **hyphae** (singular, **hypha**), whereas the reproductive stage can be more conspicuous. The mass of hyphae is a **mycelium** ([link]). It can grow on a surface, in soil or decaying material, in a liquid, or even on living tissue. Although

individual hyphae must be observed under a microscope, the mycelium of a fungus can be very large, with some species truly being “the fungus humongous.” The giant *Armillaria solidipes* (honey mushroom) is considered the largest organism on Earth, spreading across more than 2,000 acres of underground soil in eastern Oregon; it is estimated to be at least 2,400 years old.



The mycelium of the fungus *Neotestudina rosati* can be pathogenic to humans. The fungus enters through a cut or scrape and develops a mycetoma, a chronic subcutaneous infection. (credit: CDC)

Most fungal hyphae are divided into separate cells by endwalls called **septa** (singular, **septum**) ([link]**a, c**). In most phyla of fungi, tiny holes in the septa allow for the rapid flow of nutrients and small molecules from cell to cell along the hypha. They are described as perforated septa. The hyphae in bread molds (which belong to the Phylum Zygomycota) are not separated by septa. Instead, they are formed by large cells containing many nuclei, an arrangement described as **coenocytic hyphae** ([link]**b**).



Fungal hyphae may be (a) septated or (b) coenocytic (coeno- = "common"; -cytic = "cell") with many nuclei present in a single hypha. A bright field light micrograph of (c) *Phialophora richardsiae* shows septa that divide the hyphae. (credit c: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

Fungi thrive in environments that are moist and slightly acidic, and can grow with or without light. They vary in their oxygen requirement. Most fungi are **obligate aerobes**, requiring oxygen to survive. Other species, such as the Chytridiomycota that reside in the rumen of cattle, are are **obligate anaerobes**, in that they only use anaerobic respiration because oxygen will disrupt their metabolism or kill them. Yeasts are intermediate, being **faculative anaerobes**. This means that they grow best in the presence of oxygen using aerobic respiration, but can survive using anaerobic respiration when oxygen is not available. The alcohol produced from yeast fermentation is used in wine and beer production.

Nutrition

Like animals, fungi are heterotrophs; they use complex organic compounds as a source of carbon, rather than fix carbon dioxide from the atmosphere as do some bacteria and most plants. In addition, fungi do not fix nitrogen from the atmosphere. Like animals, they must obtain it from their diet. However, unlike most animals, which ingest food and then digest it internally in specialized organs, fungi perform these steps in the reverse order; digestion precedes ingestion. First, exoenzymes are transported out of the hyphae, where they process nutrients in the environment. Then, the smaller molecules produced by this external digestion are absorbed through the large surface area of the mycelium. As with animal cells, the polysaccharide of storage is glycogen, rather than starch, as found in plants.

Fungi are mostly **saprobes** (saprophyte is an equivalent term): organisms that derive nutrients from decaying organic matter. They obtain their nutrients from dead or decomposing organic matter: mainly plant material. Fungal exoenzymes are able to break down insoluble polysaccharides, such as the cellulose and lignin of dead wood, into readily absorbable glucose molecules. The carbon, nitrogen, and other elements are thus released into the environment. Because of their varied metabolic pathways, fungi fulfill an important ecological role and are being investigated as potential tools in bioremediation. For example, some species of fungi can be used to break down diesel oil and polycyclic aromatic hydrocarbons (PAHs). Other species take up heavy metals, such as cadmium and lead.

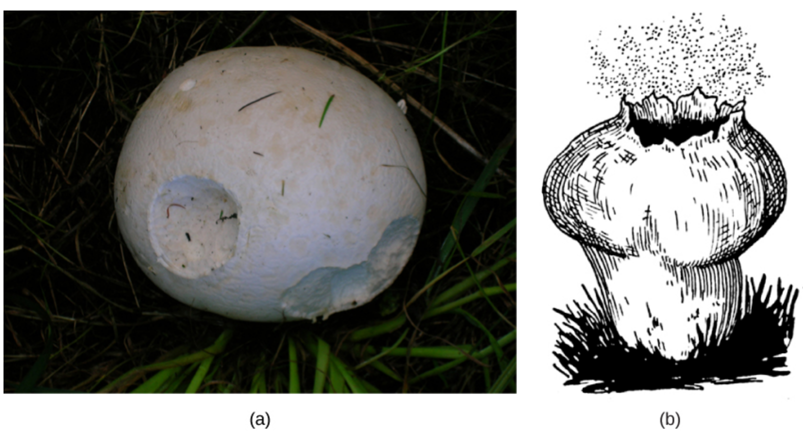
Some fungi are parasitic, infecting either plants or animals. Smut and Dutch elm disease affect plants, whereas athlete’s foot and candidiasis (thrush) are medically important fungal infections in humans. In environments poor in nitrogen, some fungi resort to predation of nematodes (small non-segmented roundworms). Species of *Arthrobotrys* fungi have a number of mechanisms to trap nematodes. One mechanism involves constricting rings within the network of hyphae. The rings swell when they touch the nematode, gripping it in a tight hold. The fungus penetrates the tissue of the worm by extending specialized hyphae called **haustoria**. Many parasitic fungi possess haustoria, as these structures penetrate the tissues of the host,

release digestive enzymes within the host's body, and absorb the digested nutrients.

Reproduction

Fungi reproduce sexually and/or asexually. Perfect fungi reproduce both sexually and asexually, while the so-called imperfect fungi reproduce only asexually (by mitosis).

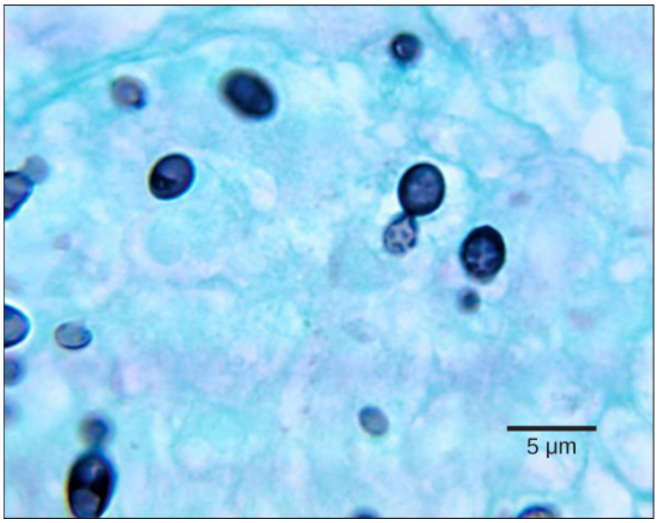
In both sexual and asexual reproduction, fungi produce spores that disperse from the parent organism by either floating on the wind or hitching a ride on an animal. Fungal spores are smaller and lighter than plant seeds. The giant puffball mushroom bursts open and releases trillions of spores. The huge number of spores released increases the likelihood of landing in an environment that will support growth ([link]).



The (a) giant puff ball mushroom releases (b) a cloud of spores when it reaches maturity. (credit a: modification of work by Roger Griffith; credit b: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

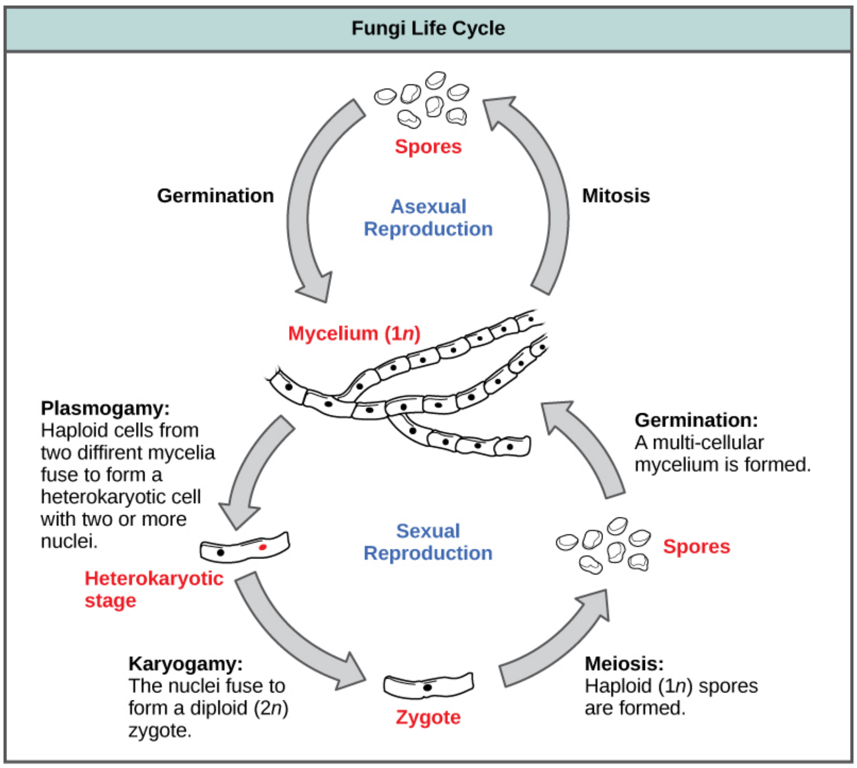
Asexual Reproduction

Fungi reproduce asexually by fragmentation, budding, or producing spores. Fragments of hyphae can grow new colonies. Somatic cells in yeast form buds. During budding (a type of cytokinesis), a bulge forms on the side of the cell, the nucleus divides mitotically, and the bud ultimately detaches itself from the mother cell ([link]).



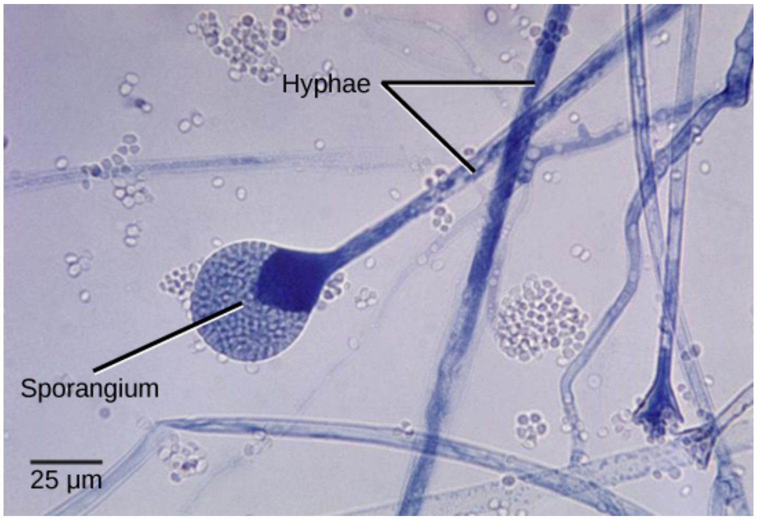
The dark cells in this bright field light micrograph are the pathogenic yeast *Histoplasma capsulatum*, seen against a backdrop of light blue tissue. Histoplasma primarily infects lungs but can spread to other tissues, causing histoplasmosis, a potentially fatal disease. (credit: modification of work by Dr. Libero Ajello, CDC; scale-bar data from Matt Russell)

The most common mode of asexual reproduction is through the formation of asexual spores, which are produced by one parent only (through mitosis) and are genetically identical to that parent ([link]). Spores allow fungi to expand their distribution and colonize new environments. They may be released from the parent thallus either outside or within a special reproductive sac called a **sporangium**.



Fungi may have both asexual and sexual stages of reproduction.

There are many types of asexual spores. Conidiospores are unicellular or multicellular spores that are released directly from the tip or side of the hypha. Other asexual spores originate in the fragmentation of a hypha to form single cells that are released as spores; some of these have a thick wall surrounding the fragment. Yet others bud off the vegetative parent cell. Sporangiospores are produced in a sporangium ([link]).



This bright field light micrograph shows the release of spores from a sporangium at the end of a hypha called a sporangiophore. The organism is a *Mucor* sp. fungus, a mold often found indoors. (credit: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

Sexual Reproduction

Sexual reproduction introduces genetic variation into a population of fungi. In fungi, sexual reproduction often occurs in response to adverse environmental conditions. During sexual reproduction, two mating types are produced. When both mating types are present in the same mycelium, it is called **homothallic**, or self-fertile. **Heterothallic** mycelia require two different, but compatible, mycelia to reproduce sexually.

Although there are many variations in fungal sexual reproduction, all include the following three stages ([link]). First, during **plasmogamy** (literally, “marriage or union of cytoplasm”), two haploid cells fuse, leading

to a dikaryotic stage where two haploid nuclei coexist in a single cell. During **karyogamy** (“nuclear marriage”), the haploid nuclei fuse to form a diploid zygote nucleus. Finally, meiosis takes place in the gametangia (singular, gametangium) organs, in which gametes of different mating types are generated. At this stage, spores are disseminated into the environment.

Note:
Link to Learning

Review the characteristics of fungi by visiting this interactive site from Wisconsin-online.

Section Summary

Fungi are eukaryotic organisms that appeared on land more than 450 million years ago. They are heterotrophs and contain neither photosynthetic pigments such as chlorophyll, nor organelles such as chloroplasts. Because fungi feed on decaying and dead matter, they are saprobes. Fungi are important decomposers that release essential elements into the environment. External enzymes digest nutrients that are absorbed by the body of the fungus, which is called a thallus. A thick cell wall made of chitin surrounds the cell. Fungi can be unicellular as yeasts, or develop a network of filaments called a mycelium, which is often described as mold. Most species multiply by asexual and sexual reproductive cycles and display an alternation of generations. Another group of fungi do not have a sexual cycle. Sexual reproduction involves plasmogamy (the fusion of the

cytoplasm), followed by karyogamy (the fusion of nuclei). Meiosis regenerates haploid individuals, resulting in haploid spores.

Review Questions

Exercise:

Problem:

Which polysaccharide is usually found in the cell wall of fungi?

a. starch
b. glycogen
c. chitin
d. cellulose

Solution:

C

Exercise:

Problem: Which of these organelles is not found in a fungal cell?

a. chloroplast
b. nucleus
c. mitochondrion
d. Golgi apparatus

Solution:

A

Exercise:

Problem:

The wall dividing individual cells in a fungal filament is called a

a. thallus
b. hypha
c. mycelium
d. septum

Solution:

D

Exercise:

Problem:

During sexual reproduction, a homothallic mycelium contains

a. all septated hyphae
b. all haploid nuclei
c. both mating types
d. none of the above

Solution:

C

Free Response

Exercise:

Problem:

What are the evolutionary advantages for an organism to reproduce both asexually and sexually?

Solution:

Asexual reproduction is fast and best under favorable conditions. Sexual reproduction allows the recombination of genetic traits and

increases the odds of developing new adaptations better suited to a changed environment.

Exercise:

Problem:

Compare plants, animals, and fungi, considering these components: cell wall, chloroplasts, plasma membrane, food source, and polysaccharide storage. Be sure to indicate fungi's similarities and differences to plants and animals.

Solution:

Animals have no cell walls; fungi have cell walls containing chitin; plants have cell walls containing cellulose. Chloroplasts are absent in both animals and fungi but are present in plants. Animal plasma membranes are stabilized with cholesterol, while fungi plasma membranes are stabilized with ergosterol, and plant plasma membranes are stabilized with phytosterols. Animals obtain N and C from food sources via internal digestion. Fungi obtain N and C from food sources via external digestion. Plants obtain organic N from the environment or through symbiotic N-fixing bacteria; they obtain C from photosynthesis. Animals and fungi store polysaccharides as glycogen, while plants store them as starch.

Glossary

coenocytic hypha
: single hypha that lacks septa and contains many nuclei

faculative anaerobes
: organisms that can perform both aerobic and anaerobic respiration and can survive in oxygen-rich and oxygen-poor environment

haustoria
: modified hyphae on many parasitic fungi that penetrate the tissues of their hosts, release digestive enzymes, and/or absorb nutrients from the

host

heterothallic
describes when only one mating type is present in an individual mycelium

homothallic
describes when both mating types are present in mycelium

hypha
fungal filament composed of one or more cells

karyogamy
fusion of nuclei

mycelium
mass of fungal hyphae

mycology
scientific study of fungi

obligate aerobes
organisms, such as humans, that must perform aerobic respiration to survive

obligate anaerobes
organisms that only perform anaerobic respiration and often cannot survive in the presence of oxygen

plasmogamy
fusion of cytoplasm

saprobe
organism that derives nutrients from decaying organic matter; also saprophyte

septa
cell wall division between hyphae

sporangium
: reproductive sac that contains spores

thallus
: vegetative body of a fungus

yeast
: general term used to describe unicellular fungi

Classifications of Fungi
By the end of this section, you will be able to:

- Classify fungi into the five major phyla
- Describe each phylum in terms of major representative species and patterns of reproduction

The kingdom Fungi contains five major phyla that were established according to their mode of sexual reproduction or using molecular data. Polyphyletic, unrelated fungi that reproduce without a sexual cycle, are placed for convenience in a sixth group called a “form phylum”. Not all mycologists agree with this scheme. Rapid advances in molecular biology and the sequencing of 18S rRNA (a part of RNA) continue to show new and different relationships between the various categories of fungi.

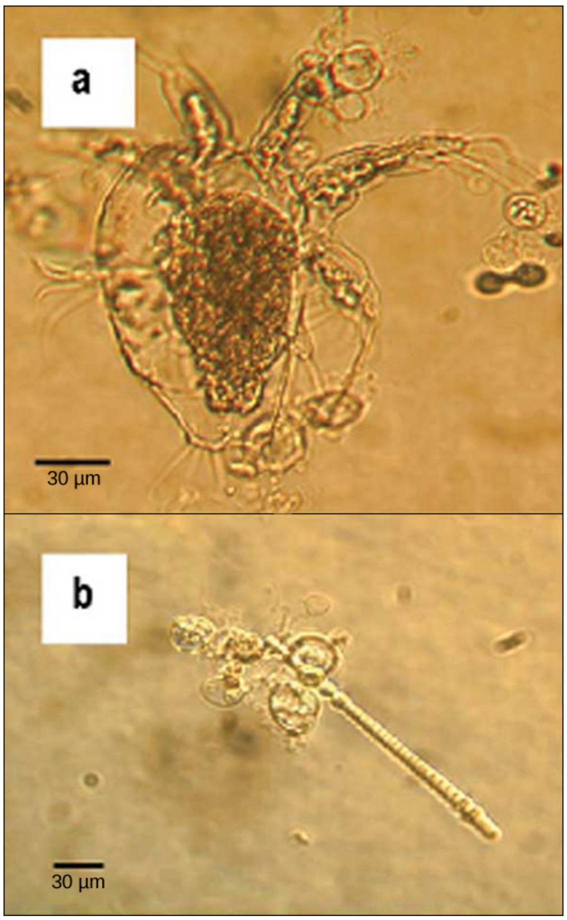
The five true phyla of fungi are the Chytridiomycota (Chytrids), the Zygomycota (conjugated fungi), the Ascomycota (sac fungi), the Basidiomycota (club fungi) and the recently described Phylum Glomeromycota. An older classification scheme grouped fungi that strictly use asexual reproduction into Deuteromycota, a group that is no longer in use.

Note: “-mycota” is used to designate a phylum while “-mycetes” formally denotes a class or is used informally to refer to all members of the phylum.

Chytridiomycota: The Chytrids

The only class in the Phylum Chytridiomycota is the **Chytridiomycetes**. The chytrids are the simplest and most primitive Eumycota, or true fungi. The evolutionary record shows that the first recognizable chytrids appeared during the late pre-Cambrian period, more than 500 million years ago. Like all fungi, chytrids have chitin in their cell walls, but one group of chytrids has both cellulose and chitin in the cell wall. Most chytrids are unicellular; a few form multicellular organisms and hyphae, which have no septa between cells (coenocytic). They produce gametes and diploid zoospores that swim with the help of a single flagellum.

The ecological habitat and cell structure of chytrids have much in common with protists. Chytrids usually live in aquatic environments, although some species live on land. Some species thrive as parasites on plants, insects, or amphibians ([link]), while others are saprobes. The chytrid species *Allomyces* is well characterized as an experimental organism. Its reproductive cycle includes both asexual and sexual phases. *Allomyces* produces diploid or haploid flagellated zoospores in a sporangium.



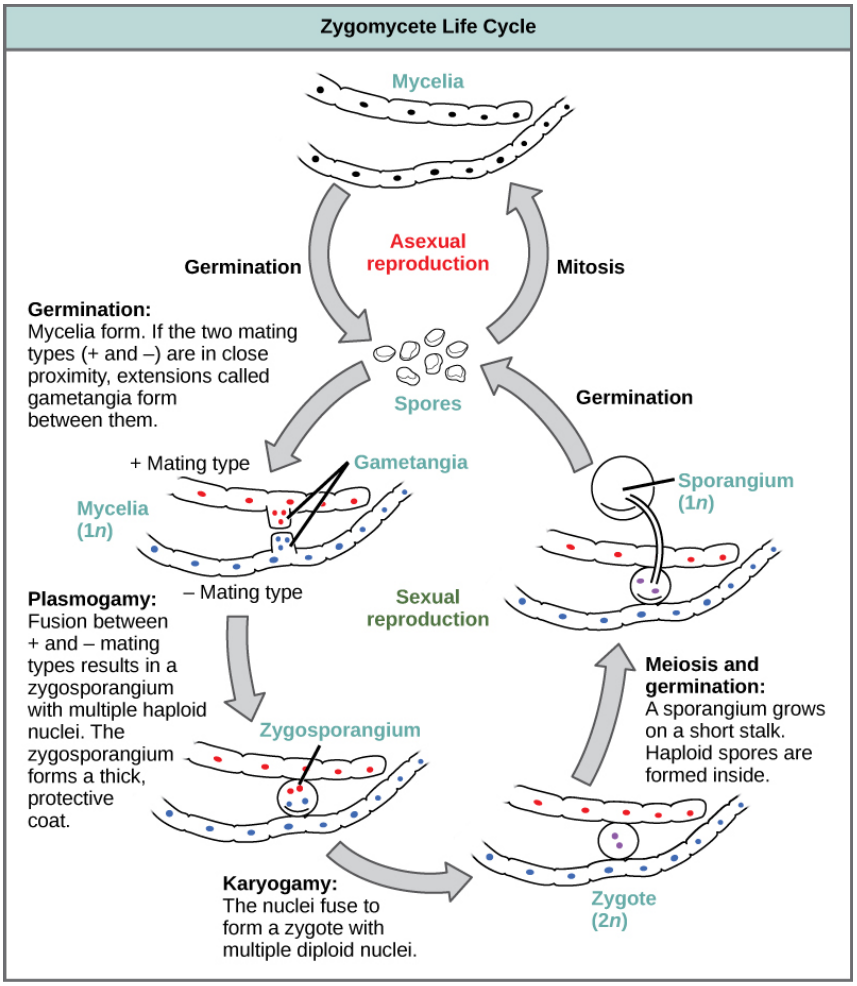
The chytrid *Batrachochytrium dendrobatidis* is seen in these light micrographs as transparent spheres growing on (a) a freshwater arthropod and (b) algae. This chytrid

causes skin diseases in many species of amphibians, resulting in species decline and extinction. (credit: modification of work by Johnson ML, Speare R., CDC)

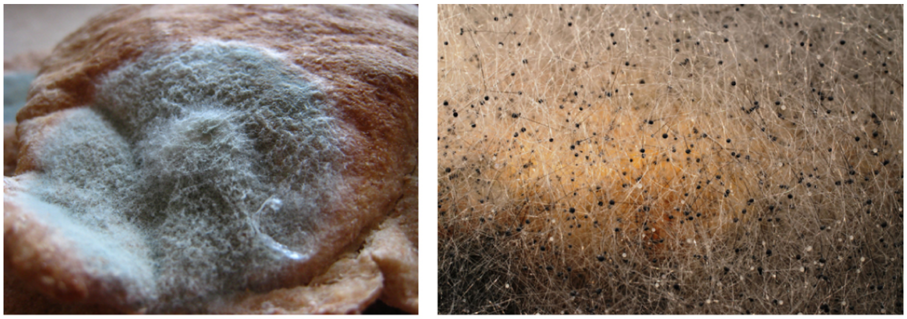
Zygomycota: The Conjugated Fungi

The zygomycetes are a relatively small group of fungi belonging to the Phylum **Zygomycota**. They include the familiar bread mold, *Rhizopus stolonifer*, which rapidly propagates on the surfaces of breads, fruits, and vegetables. Most species are saprobes, living off decaying organic material; a few are parasites, particularly of insects. Zygomycetes play a considerable commercial role. The metabolic products of other species of *Rhizopus* are intermediates in the synthesis of semi-synthetic steroid hormones.

Zygomycetes have a thallus of coenocytic hyphae in which the nuclei are haploid when the organism is in the vegetative stage. The fungi usually reproduce asexually by producing sporangiospores ([link]). The black tips of bread mold are the swollen sporangia packed with black spores ([link]). When spores land on a suitable substrate, they germinate and produce a new mycelium. Sexual reproduction starts when conditions become unfavorable. Two opposing mating strains (type + and type –) must be in close proximity for gametangia from the hyphae to be produced and fuse, leading to karyogamy. The developing diploid **zygospores** have thick coats that protect them from desiccation and other hazards. They may remain dormant until environmental conditions are favorable. When the zygospore germinates, it undergoes meiosis and produces haploid spores, which will, in turn, grow into a new organism. This form of sexual reproduction in fungi is called conjugation (although it differs markedly from conjugation in bacteria and protists), giving rise to the name “conjugated fungi”.



Zygomycetes have asexual and asexual life cycles. In the sexual life cycle, plus and minus mating types conjugate to form a zygosporangium.



(a) (b)

Sporangia grow at the end of stalks, which appear as (a) white fuzz seen on this bread mold, *Rhizopus stolonifer*. The (b) tips of bread mold are the spore-containing sporangia. (credit b: modification of work by "polandeze"/Flickr)

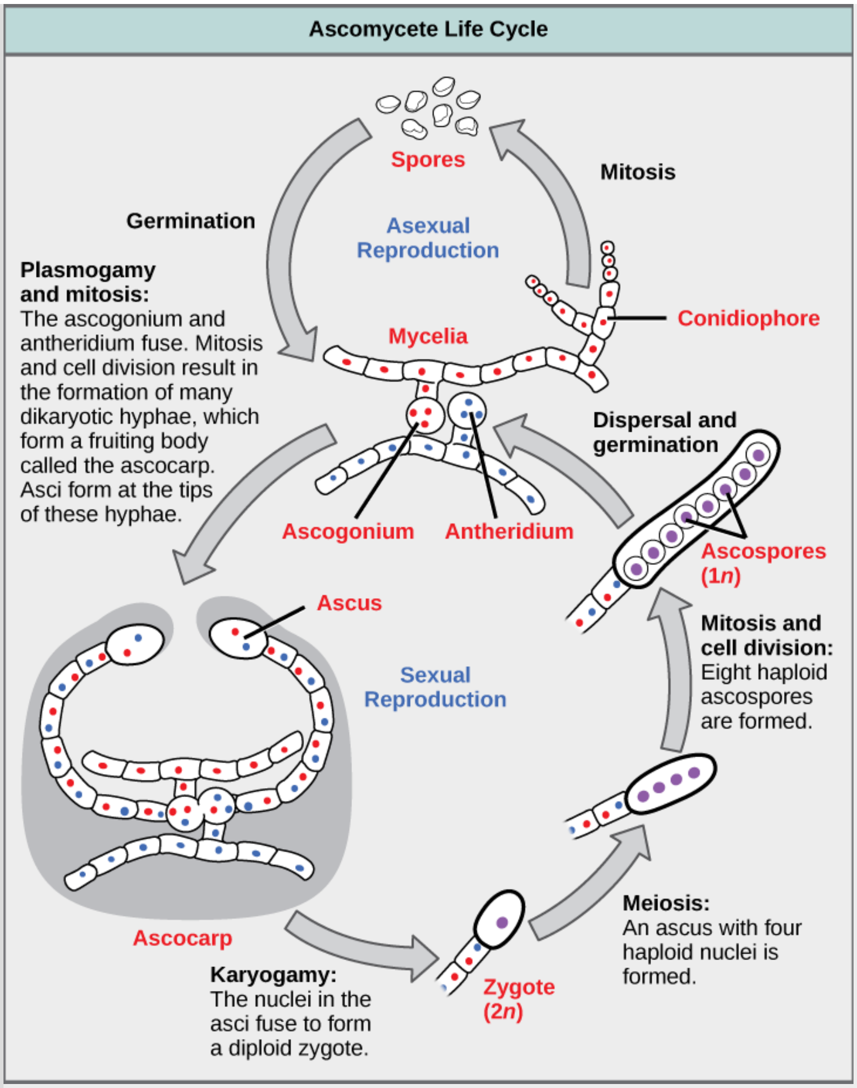
Ascomycota: The Sac Fungi

The majority of known fungi belong to the Phylum **Ascomycota**, which is characterized by the formation of an **ascus** (plural, asci), a sac-like structure that contains haploid ascospores. Many ascomycetes are of commercial importance. Some play a beneficial role, such as the yeasts used in baking, brewing, and wine fermentation, plus truffles and morels, which are held as gourmet delicacies. *Aspergillus oryzae* is used in the fermentation of rice to produce sake. Other ascomycetes parasitize plants and animals, including humans. For example, fungal pneumonia poses a significant threat to AIDS patients who have a compromised immune system. Ascomycetes not only infest and destroy crops directly; they also produce poisonous secondary metabolites that make crops unfit for consumption. Filamentous ascomycetes produce hyphae divided by perforated septa, allowing streaming of cytoplasm from one cell to the other. Conidia and asci, which are used respectively for asexual and sexual reproductions, are usually separated from the vegetative hyphae by blocked (non-perforated) septa.

Asexual reproduction is frequent and involves the production of conidiophores that release haploid conidiospores ([link]). Sexual

reproduction starts with the development of special hyphae from either one of two types of mating strains ([link]). The “male” strain produces an antheridium and the “female” strain develops an ascogonium. At fertilization, the antheridium and the ascogonium combine in plasmogamy without nuclear fusion. Special ascogenous hyphae arise, in which pairs of nuclei migrate: one from the “male” strain and one from the “female” strain. In each ascus, two or more haploid ascospores fuse their nuclei in karyogamy. During sexual reproduction, thousands of asci fill a fruiting body called the **ascocarp**. The diploid nucleus gives rise to haploid nuclei by meiosis. The ascospores are then released, germinate, and form hyphae that are disseminated in the environment and start new mycelia ([link]).

Note:
Art Connection

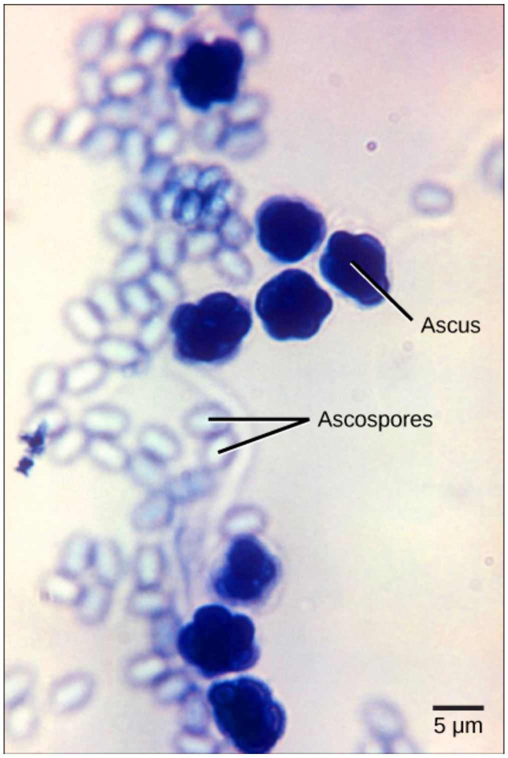


The lifecycle of an ascomycete is characterized by the production of asci during the sexual phase. The haploid phase is the predominant phase of the life cycle.

Which of the following statements is true?

a. A dikaryotic ascus that forms in the ascocarp undergoes karyogamy, meiosis, and mitosis to form eight ascospores.
b. A diploid ascus that forms in the ascocarp undergoes karyogamy, meiosis, and mitosis to form eight ascospores.

c. A haploid zygote that forms in the ascocarp undergoes karyogamy, meiosis, and mitosis to form eight ascospores.
d. A dikaryotic ascus that forms in the ascocarp undergoes plasmogamy, meiosis, and mitosis to form eight ascospores.



The bright field light micrograph shows ascospores being released from asci in the fungus *Talaromyces flavus* var. *flavus*. (credit: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

Basidiomycota: The Club Fungi

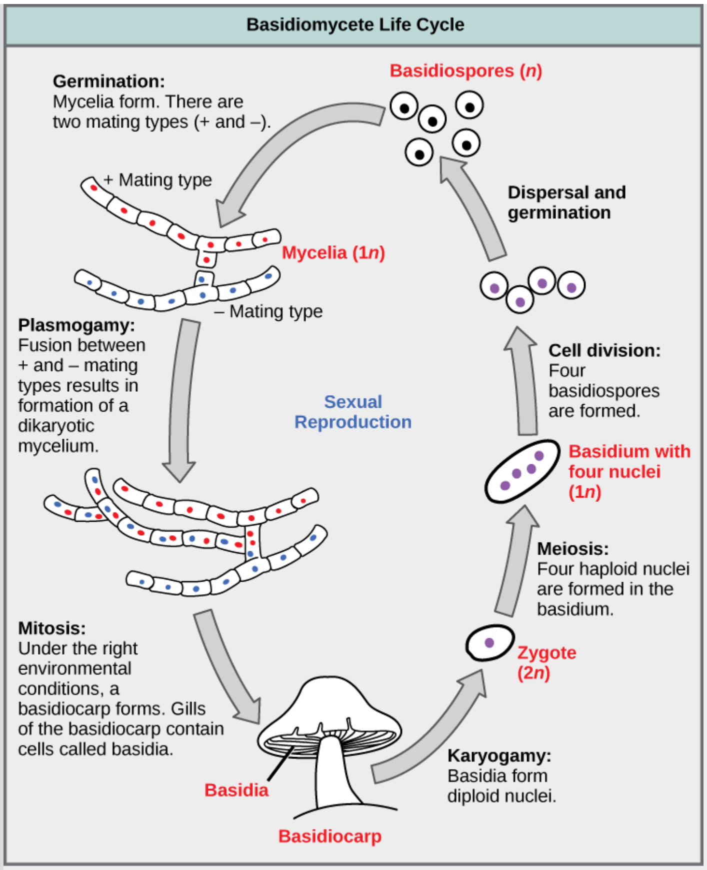
The fungi in the Phylum **Basidiomycota** are easily recognizable under a light microscope by their club-shaped fruiting bodies called **basidia** (singular, **basidium**), which are the swollen terminal cell of a hypha. The basidia, which are the reproductive organs of these fungi, are often contained within the familiar mushroom, commonly seen in fields after rain, on the supermarket shelves, and growing on your lawn ([link]). These mushroom-producing basidiomyces are sometimes referred to as "gill fungi" because of the presence of gill-like structures on the underside of the cap. The "gills" are actually compacted hyphae on which the basidia are borne. This group also includes shelf fungus, which cling to the bark of trees like small shelves. In addition, the basidiomycota includes smuts and rusts, which are important plant pathogens; toadstools, and shelf fungi stacked on tree trunks. Most edible fungi belong to the Phylum Basidiomycota; however, some basidiomycetes produce deadly toxins. For example, *Cryptococcus neoformans* causes severe respiratory illness.

The fruiting bodies of a basidiomycete form a ring in a meadow, commonly called "fairy ring." The best-known fairy ring fungus has the scientific name *Marasmius oreades*. The body of this

fungus, its mycelium, is underground and grows outward in a circle. As it grows, the mycelium depletes the soil of nitrogen, causing the mycelia to grow away from the center and leading to the “fairy ring” of fruiting bodies where there is adequate soil nitrogen. (Credit: "Cropcircles"/Wikipedia Commons)]

The lifecycle of basidiomycetes includes alternation of generations ([link]). Spores are generally produced through sexual reproduction, rather than asexual reproduction. The club-shaped basidium carries spores called basidiospores. In the basidium, nuclei of two different mating strains fuse (karyogamy), giving rise to a diploid zygote that then undergoes meiosis. The haploid nuclei migrate into basidiospores, which germinate and generate monokaryotic hyphae. The mycelium that results is called a primary mycelium. Mycelia of different mating strains can combine and produce a secondary mycelium that contains haploid nuclei of two different mating strains. This is the dikaryotic stage of the basidiomyces lifecyle and and it is the dominant stage. Eventually, the secondary mycelium generates a **basidiocarp**, which is a fruiting body that protrudes from the ground—this is what we think of as a mushroom. The basidiocarp bears the developing basidia on the gills under its cap.

Note:
Art Connection



The lifecycle of a basidiomycete alternates generation with a prolonged stage in which two nuclei (dikaryon) are present in the hyphae.

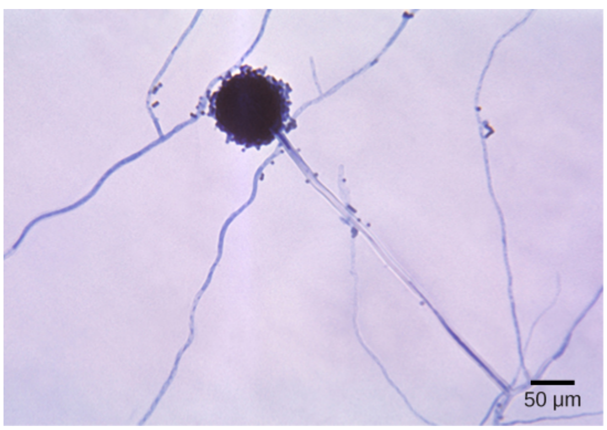
Which of the following statements is true?

a. A basidium is the fruiting body of a mushroom-producing fungus, and it forms four basidiocarps.
b. The result of the plasmogamy step is four basidiospores.
c. Karyogamy results directly in the formation of mycelia.
d. A basidiocarp is the fruiting body of a mushroom-producing fungus.

Asexual Ascomycota and Basidiomycota

Imperfect fungi—those that do not display a sexual phase—use to be classified in the form phylum **Deuteromycota**, , a classification group no longer used in the present, ever-developing classification of organisms. While Deuteromycota use to be a classification group, recent moleclular analysis has shown that the members classified in this group belong to the Ascomycota or the Basidiomycota classifications. Since they do not possess the sexual structures that are used to classify other fungi, they are less well described in comparison to other members. Most members live on land, with a few aquatic exceptions. They form visible mycelia with a fuzzy appearance and are commonly known as **mold**.

Reproduction of the fungi in this group is strictly asexual and occurs mostly by production of asexual conidiospores ([link]). Some hyphae may recombine and form heterokaryotic hyphae. Genetic recombination is known to take place between the different nuclei.



Aspergillus niger is an asexually reproducing fungus (phylum Ascomycota) commonly found as a food contaminant. The spherical structure in this light micrograph is a conidiophore.

(credit: modification of work by Dr. Lucille Georg, CDC; scale-bar data from Matt Russell)

The fungi in this group have a large impact on everyday human life. The food industry relies on them for ripening some cheeses. The blue veins in Roquefort cheese and the white crust on Camembert are the result of fungal growth. The antibiotic penicillin was originally discovered on an overgrown Petri plate, on which a colony of *Penicillium* fungi killed the bacterial growth surrounding it. Other fungi in this group cause serious diseases, either directly as parasites (which infect both plants and humans), or as producers of potent toxic compounds, as seen in the aflatoxins released by fungi of the genus *Aspergillus*.

Glomeromycota

The **Glomeromycota** is a newly established phylum which comprises about 230 species that all live in close association with the roots of trees. Fossil records indicate that trees and their root symbionts share a long evolutionary history. It appears that all members of this family form **arbuscular mycorrhizae**: the hyphae interact with the root cells forming a mutually beneficial association where the plants supply the carbon source and energy in the form of carbohydrates to the fungus, and the fungus supplies essential minerals from the soil to the plant.

The glomeromycetes do not reproduce sexually and do not survive without the presence of plant roots. Although they have coenocytic hyphae like the zygomycetes, they do not form zygospores. DNA analysis shows that all glomeromycetes probably descended from a common ancestor, making them a monophyletic lineage.

Section Summary

Chytridiomycota (chytrids) are considered the most primitive group of fungi. They are mostly aquatic, and their gametes are the only fungal cells known to have flagella. They reproduce both sexually and asexually; the asexual spores are called zoospores. Zygomycota (conjugated fungi) produce non-septated hyphae with many nuclei. Their hyphae fuse during sexual reproduction to produce a zygospore in a zygosporangium. Ascomycota (sac fungi) form spores in sacs called asci during sexual reproduction. Asexual reproduction is their most common form of reproduction. Basidiomycota (club fungi) produce showy fruiting bodies that contain basidia in the form of clubs. Spores are stored in the basidia. Most familiar mushrooms belong to this division. Fungi that have no known sexual cycle were classified in the form phylum Deuteromycota, which the present classification puts in the phyla Ascomycota and Basidiomycota. Glomeromycota form tight associations (called mycorrhizae) with the roots of plants.

Art Connections

Exercise:

Problem: [link] Which of the following statements is true?

a. A dikaryotic ascus that forms in the ascocarp undergoes karyogamy, meiosis, and mitosis to form eight ascospores.
b. A diploid ascus that forms in the ascocarp undergoes karyogamy, meiosis, and mitosis to form eight ascospores.
c. A haploid zygote that forms in the ascocarp undergoes karyogamy, meiosis, and mitosis to form eight ascospores.
d. A dikaryotic ascus that forms in the ascocarp undergoes plasmogamy, meiosis, and mitosis to form eight ascospores.

Solution:

[link] A

Exercise:

Problem: [link] Which of the following statements is true?

a. A basidium is the fruiting body of a mushroom-producing fungus, and it forms four basidiocarps.
b. The result of the plasmogamy step is four basidiospores.
c. Karyogamy results directly in the formation of mycelia.
d. A basidiocarp is the fruiting body of a mushroom-producing fungus.

Solution:

[link] D

Review Questions

Exercise:

Problem: The most primitive phylum of fungi is the ________.

a. Chytridiomycota
b. Zygomycota
c. Glomeromycota
d. Ascomycota

Solution:

A

Exercise:

Problem:

Members of which phylum produce a club-shaped structure that contains spores?

a. Chytridiomycota
b. Basidiomycota
c. Glomeromycota
d. Ascomycota

Solution:

B

Exercise:

Problem:

Members of which phylum establish a successful symbiotic relationship with the roots of trees?

a. Ascomycota
b. Deuteromycota
c. Basidiomycota
d. Glomeromycota

Solution:

D

Exercise:

Problem:

The fungi that do not reproduce sexually use to be classified as ________.

a. Ascomycota
b. Deuteromycota
c. Basidiomycota
d. Glomeromycota

Solution:

B

Free Response

Exercise:

Problem:

What is the advantage for a basidiomycete to produce a showy and fleshy fruiting body?

Solution:

By ingesting spores and disseminating them in the environment as waste, animals act as agents of dispersal. The benefit to the fungus outweighs the cost of producing fleshy fruiting bodies.

Exercise:

Problem:

For each of the four groups of perfect fungi (Chytridiomycota, Zygomycota, Ascomycota, and Basidiomycota), compare the body structure and features, and provide an example.

Solution:

Chytridiomycota (Chytrids) may have a unicellular or multicellular body structure; some are aquatic with motile spores with flagella; an example is the *Allomyces*. Zygomycota (conjugated fungi) have a multicellular body structure; features include zygospores and presence in soil; examples are bread and fruit molds. Ascomycota (sac fungi) may have unicellular or multicellular body structure; a feature is sexual spores in sacs (asci); examples include the yeasts used in bread, wine, and beer production. Basidiomycota (club fungi) have multicellular bodies; features includes sexual spores in the basidiocarp (mushroom) and that they are mostly decomposers; mushroom-producing fungi are an example.

Glossary

Arbuscular mycorrhizae
: mycorrhizae commonly involving Glomeromycetes in which the fungal hyphae penetrate the cell walls of the plant root cells (but not the cell membranes)

ascocarp
: fruiting body of ascomycetes

Ascomycota
: (also, sac fungi) phylum of fungi that store spores in a sac called ascus

basidiocarp
: fruiting body that protrudes from the ground and bears the basidia

Basidiomycota
: (also, club fungi) phylum of fungi that produce club-shaped structures (basidia) that contain spores

basidium
: club-shaped fruiting body of basidiomycetes

Chytridiomycota
: (also, chytrids) primitive phylum of fungi that live in water and produce gametes with flagella

Deuteromycota
: former form phylum of fungi that do not have a known sexual reproductive cycle (presently members of two phyla: Ascomycota and Basidiomycota)

Ectomycorrhizae
: mycorrhizae in which the fungal hyphae do not penetrate the root cells of the plant

Glomeromycota

phylum of fungi that form symbiotic relationships with the roots of trees

mold
: tangle of visible mycelia with a fuzzy appearance

Zygomycota
: (also, conjugated fungi) phylum of fungi that form a zygote contained in a zygospore

zygospore
: structure with thick cell wall that contains the zygote in zygomycetes

Ecology of Fungi

By the end of this section, you will be able to:

- Describe the role of fungi in the ecosystem
- Describe mutualistic relationships of fungi with plant roots and photosynthetic organisms
- Describe the beneficial relationship between some fungi and insects

Fungi play a crucial role in the balance of ecosystems. They colonize most habitats on Earth, preferring dark, moist conditions. They can thrive in seemingly hostile environments, such as the tundra, thanks to a most successful symbiosis with photosynthetic organisms like algae to produce lichens. Fungi are not obvious in the way large animals or tall trees appear. Yet, like bacteria, they are the major decomposers of nature. With their versatile metabolism, fungi break down organic matter, which would not otherwise be recycled.

Habitats

Although fungi are primarily associated with humid and cool environments that provide a supply of organic matter, they colonize a surprising diversity of habitats, from seawater to human skin and mucous membranes. Chytrids are found primarily in aquatic environments. Other fungi, such as *Coccidioides immitis*, which causes pneumonia when its spores are inhaled, thrive in the dry and sandy soil of the southwestern United States. Fungi that parasitize coral reefs live in the ocean. However, most members of the Kingdom Fungi grow on the forest floor, where the dark and damp environment is rich in decaying debris from plants and animals. In these environments, fungi play a major role as decomposers and recyclers, making it possible for members of the other kingdoms to be supplied with nutrients and live.

Decomposers and Recyclers

The food web would be incomplete without organisms that decompose organic matter ([link]). Some elements—such as nitrogen and phosphorus—are required in large quantities by biological systems, and yet are not

abundant in the environment. The action of fungi releases these elements from decaying matter, making them available to other living organisms. Trace elements present in low amounts in many habitats are essential for growth, and would remain tied up in rotting organic matter if fungi and bacteria did not return them to the environment via their metabolic activity.

Fungi are an important part of ecosystem nutrient cycles. These bracket fungi growing on the side of a tree are the fruiting structures of a basidiomycete. They receive their nutrients through their hyphae, which invade and decay the tree trunk. (credit: Cory Zanker)

The ability of fungi to degrade many large and insoluble molecules is due to their mode of nutrition. As seen earlier, digestion precedes ingestion. Fungi produce a variety of exoenzymes to digest nutrients. The enzymes are either released into the substrate or remain bound to the outside of the fungal cell wall. Large molecules are broken down into small molecules, which are transported into the cell by a system of protein carriers embedded in the cell membrane. Because the movement of small molecules and enzymes is

dependent on the presence of water, active growth depends on a relatively high percentage of moisture in the environment.

As saprobes, fungi help maintain a sustainable ecosystem for the animals and plants that share the same habitat. In addition to replenishing the environment with nutrients, fungi interact directly with other organisms in beneficial, and sometimes damaging, ways ([link]).

Shelf fungi, so called because they grow on trees in a stack, attack and digest the trunk or branches of a tree. While some shelf fungi are found only on dead trees, others can parasitize living trees and cause eventual death, so they are considered serious tree pathogens. (credit: Cory Zanker)

Mutualistic Relationships

Symbiosis is the ecological interaction between two organisms that live together. The definition does not describe the quality of the interaction. When both members of the association benefit, the symbiotic relationship is

called mutualistic. Fungi form mutualistic associations with many types of organisms, including cyanobacteria, algae, plants, and animals.

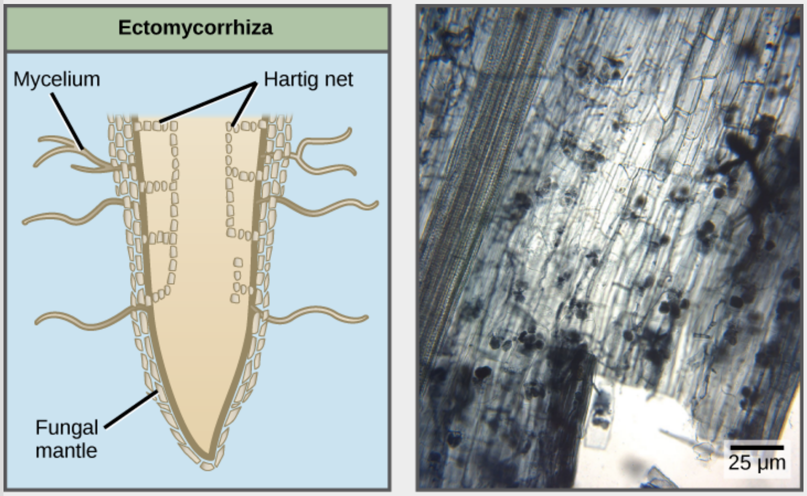
Fungus/Plant Mutualism

One of the most remarkable associations between fungi and plants is the establishment of mycorrhizae. **Mycorrhiza**, which comes from the Greek words *myco* meaning fungus and *rhizo* meaning root, refers to the association between vascular plant roots and their symbiotic fungi. Somewhere between 80 and 90 percent of all plant species have mycorrhizal partners. In a mycorrhizal association, the fungal mycelia use their extensive network of hyphae and large surface area in contact with the soil to channel water and minerals from the soil into the plant. In exchange, the plant supplies the products of photosynthesis to fuel the metabolism of the fungus.

There are a number of types of mycorrhizae. **Ectomycorrhizae** ("outside" mycorrhiza) depend on fungi enveloping the roots in a sheath (called a mantle) and a Hartig net of hyphae that extends into the roots between cells ([link]). The fungal partner can belong to the Ascomycota, Basidiomycota or Zygomycota. In a second type, the Glomeromycete fungi form vesicular–arbuscular interactions with **arbuscular mycorrhiza** (sometimes called endomycorrhizae). In these mycorrhiza, the fungi form arbuscules that penetrate root cells and are the site of the metabolic exchanges between the fungus and the host plant ([link] and [link]). The arbuscules (from the Latin for little trees) have a shrub-like appearance. Orchids rely on a third type of mycorrhiza. Orchids are epiphytes that form small seeds without much storage to sustain germination and growth. Their seeds will not germinate without a mycorrhizal partner (usually a Basidiomycete). After nutrients in the seed are depleted, fungal symbionts support the growth of the orchid by providing necessary carbohydrates and minerals. Some orchids continue to be mycorrhizal throughout their lifecycle.

Note:

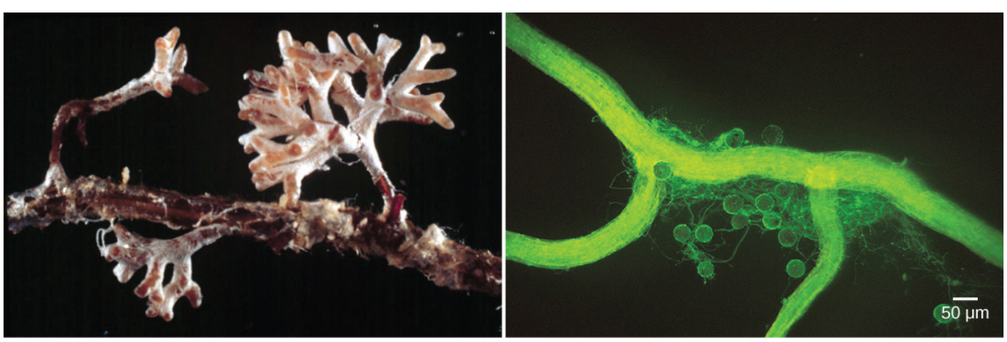
Art Connection



(a) (b)

(a) Ectomycorrhiza and (b) arbuscular mycorrhiza have different mechanisms for interacting with the roots of plants. (credit b: MS Turmel, University of Manitoba, Plant Science Department)

If symbiotic fungi are absent from the soil, what impact do you think this would have on plant growth?



(a) (b)

The (a) infection of *Pinus radiata* (Monterey pine) roots by the hyphae of *Amanita muscaria* (fly amanita) causes the pine tree to produce many small, branched rootlets.

The *Amanita* hyphae cover these small roots with a white mantle. (b) Spores (round bodies) and hyphae (thread-like structures) are evident in this light micrograph of an arbuscular mycorrhiza between a fungus and the root of a corn plant. (credit a: modification of work by Randy Molina, USDA; credit b: modification of work by Sara Wright, USDA-ARS; scale-bar data from Matt Russell)

Other examples of fungus–plant mutualism include the endophytes: fungi that live inside tissue without damaging the host plant. Endophytes release toxins that repel herbivores, or confer resistance to environmental stress factors, such as infection by microorganisms, drought, or heavy metals in soil.

Note:
Evolution Connection
Coevolution of Land Plants and Mycorrhizae
Mycorrhizae are the mutually beneficial symbiotic association between roots of vascular plants and fungi. A well-accepted theory proposes that fungi were instrumental in the evolution of the root system in plants and contributed to the success of Angiosperms. The bryophytes (mosses and liverworts), which are considered the most primitive plants and the first to survive on dry land, do not have a true root system; some have vesicular–arbuscular mycorrhizae and some do not. They depend on a simple rhizoid (an underground organ) and cannot survive in dry areas. True roots appeared in vascular plants. Vascular plants that developed a system of thin extensions from the rhizoids (found in mosses) are thought to have had a selective advantage because they had a greater surface area of contact with the fungal partners than the mosses and liverworts, thus availing themselves of more nutrients in the ground.
Fossil records indicate that fungi preceded plants on dry land. The first association between fungi and photosynthetic organisms on land involved moss-like plants and endophytes. These early associations developed

before roots appeared in plants. Slowly, the benefits of the endophyte and rhizoid interactions for both partners led to present-day mycorrhizae; up to about 90 percent of today’s vascular plants have associations with fungi in their rhizosphere. The fungi involved in mycorrhizae display many characteristics of primitive fungi; they produce simple spores, show little diversification, do not have a sexual reproductive cycle, and cannot live outside of a mycorrhizal association. The plants benefited from the association because mycorrhizae allowed them to move into new habitats because of increased uptake of nutrients, and this gave them a selective advantage over plants that did not establish symbiotic relationships.

Lichens

Lichens display a range of colors and textures ([link]) and can survive in the most unusual and hostile habitats. They cover rocks, gravestones, tree bark, and the ground in the tundra where plant roots cannot penetrate. Lichens can survive extended periods of drought, when they become completely desiccated, and then rapidly become active once water is available again.

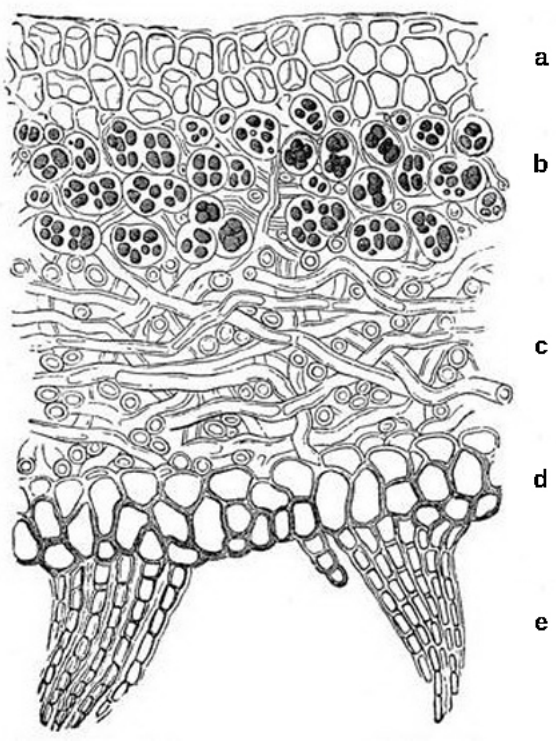
Note:
Link to Learning

Explore the world of lichens using this site from Oregon State University.

Lichens have many forms. They may be (a) crust-like, (b) hair-like, or (c) leaf-like. (credit a: modification of work by Jo Naylor; credit b: modification of work by "djpmapleferryman"/Flickr; credit c: modification of work by Cory Zanker)

(a) (b) (c)

Lichens are not a single organism, but rather an example of a mutualism, in which a fungus (usually a member of the Ascomycota or Basidiomycota phyla) lives in close contact with a photosynthetic organism (a eukaryotic alga or a prokaryotic cyanobacterium) ([link]). Generally, neither the fungus nor the photosynthetic organism can survive alone outside of the symbiotic relationship. The body of a lichen, referred to as a thallus, is formed of hyphae wrapped around the photosynthetic partner. The photosynthetic organism provides carbon and energy in the form of carbohydrates. Some cyanobacteria fix nitrogen from the atmosphere, contributing nitrogenous compounds to the association. In return, the fungus supplies minerals and protection from dryness and excessive light by encasing the algae in its mycelium. The fungus also attaches the symbiotic organism to the substrate.



This cross-section of a lichen thallus shows the (a) upper cortex of fungal hyphae, which provides protection; the (b) algal zone where photosynthesis occurs, the (c) medulla of fungal hyphae, and the (d) lower cortex, which also provides protection and may have (e) rhizines to anchor the thallus to the substrate.

The thallus of lichens grows very slowly, expanding its diameter a few millimeters per year. Both the fungus and the alga participate in the formation of dispersal units for reproduction. Lichens produce **soredia**, clusters of algal cells surrounded by mycelia. Soredia are dispersed by wind and water and form new lichens.

Lichens are extremely sensitive to air pollution, especially to abnormal levels of nitrogen and sulfur. The U.S. Forest Service and National Park Service can monitor air quality by measuring the relative abundance and health of the lichen population in an area. Lichens fulfill many ecological roles. Caribou and reindeer eat lichens, and they provide cover for small invertebrates that hide in the mycelium. In the production of textiles, weavers used lichens to dye wool for many centuries until the advent of synthetic dyes.

Note:
Link to Learning

Lichens are used to monitor the quality of air. Read more on this site from the United States Forest Service.

Fungus/Animal Mutualism

Fungi have evolved mutualisms with numerous insects in Phylum Arthropoda: jointed, legged invertebrates. Arthropods depend on the fungus for protection from predators and pathogens, while the fungus obtains nutrients and a way to disseminate spores into new environments. The association between species of Basidiomycota and scale insects is one example. The fungal mycelium covers and protects the insect colonies. The scale insects foster a flow of nutrients from the parasitized plant to the fungus. In a second example, leaf-cutting ants of Central and South America literally farm fungi. They cut disks of leaves from plants and pile them up in gardens ([link]). Fungi are cultivated in these disk gardens,

digesting the cellulose in the leaves that the ants cannot break down. Once smaller sugar molecules are produced and consumed by the fungi, the fungi in turn become a meal for the ants. The insects also patrol their garden, preying on competing fungi. Both ants and fungi benefit from the association. The fungus receives a steady supply of leaves and freedom from competition, while the ants feed on the fungi they cultivate.

A leaf cutting ant transports a leaf that will feed a farmed fungus. (credit: Scott Bauer, USDA-ARS)

Fungivores

Animal dispersal is important for some fungi because an animal may carry spores considerable distances from the source. Fungal spores are rarely completely degraded in the gastrointestinal tract of an animal, and many are able to germinate when they are passed in the feces. Some dung fungi actually require passage through the digestive system of herbivores to complete their lifecycle. The black truffle—a prized gourmet delicacy—is the fruiting body of an underground mushroom. Almost all truffles are

ectomycorrhizal, and are usually found in close association with trees. Animals eat truffles and disperse the spores. In Italy and France, truffle hunters use female pigs to sniff out truffles. Female pigs are attracted to truffles because the fungus releases a volatile compound closely related to a pheromone produced by male pigs.

Section Summary

Fungi have colonized nearly all environments on Earth, but are frequently found in cool, dark, moist places with a supply of decaying material. Fungi are saprobes that decompose organic matter. Many successful mutualistic relationships involve a fungus and another organism. Many fungi establish complex mycorrhizal associations with the roots of plants. Some ants farm fungi as a supply of food. Lichens are a symbiotic relationship between a fungus and a photosynthetic organism, usually an alga or cyanobacterium. The photosynthetic organism provides energy derived from light and carbohydrates, while the fungus supplies minerals and protection. Some animals that consume fungi help disseminate spores over long distances.

Art Connections

Exercise:

Problem:

[link] If symbiotic fungi are absent from the soil, what impact do you think this would have on plant growth?

Solution:

[link] Without mycorrhiza, plants cannot absorb adequate nutrients, which stunts their growth. Addition of fungal spores to sterile soil can alleviate this problem.

Review Questions

Exercise:

Problem:

What term describes the close association of a fungus with the root of a tree?

a. a rhizoid
b. a lichen
c. a mycorrhiza
d. an endophyte

Solution:

C

Exercise:

Problem: Why are fungi important decomposers?

a. They produce many spores.
b. They can grow in many different environments.
c. They produce mycelia.
d. They recycle carbon and inorganic minerals by the process of decomposition.

Solution:

D

Free Response

Exercise:

Problem:

Why does protection from light actually benefit the photosynthetic partner in lichens?

Solution:

Protection from excess light that may bleach photosynthetic pigments allows the photosynthetic partner to survive in environments unfavorable to plants.

Glossary

arbuscular mycorrhiza
: mycorrhizal association in which the fungal hyphae enter the root cells and form extensive networks

ectomycorrhiza
: mycorrhizal fungi that surround the roots with a mantle and have a Hartig net that extends into the roots between cells

lichen
: close association of a fungus with a photosynthetic alga or bacterium that benefits both partners

mycorrhiza
: mutualistic association between fungi and vascular plant roots

soredia
: clusters of algal cells and mycelia that allow lichens to propagate

Fungal Parasites and Pathogens

By the end of this section, you will be able to:

- Describe fungal parasites and pathogens of plants
- Describe the different types of fungal infections in humans
- Explain why antifungal therapy is hampered by the similarity between fungal and animal cells

Parasitism describes a symbiotic relationship in which one member of the association benefits at the expense of the other. Both parasites and pathogens harm the host; however, the pathogen causes a disease, whereas the parasite usually does not. **Commensalism** occurs when one member benefits without affecting the other.

Plant Parasites and Pathogens

The production of sufficient good-quality crops is essential to human existence. Plant diseases have ruined crops, bringing widespread famine. Many plant pathogens are fungi that cause tissue decay and eventual death of the host ([link]). In addition to destroying plant tissue directly, some plant pathogens spoil crops by producing potent toxins. Fungi are also responsible for food spoilage and the rotting of stored crops. For example, the fungus *Claviceps purpurea* causes ergot, a disease of cereal crops (especially of rye). Although the fungus reduces the yield of cereals, the effects of the ergot's alkaloid toxins on humans and animals are of much greater significance. In animals, the disease is referred to as ergotism. The most common signs and symptoms are convulsions, hallucination, gangrene, and loss of milk in cattle. The active ingredient of ergot is lysergic acid, which is a precursor of the drug LSD. Smuts, rusts, and powdery or downy mildew are other examples of common fungal pathogens that affect crops.

Some fungal pathogens include (a) green mold on grapefruit, (b) powdery mildew on a zinnia, (c) stem rust on a sheaf of barley, and (d) grey rot on grapes. In wet conditions *Botrytis cinerea*, the fungus that

causes grey rot, can destroy a grape crop. However, controlled infection of grapes by *Botrytis* results in noble rot, a condition that produces strong and much-prized dessert wines. (credit a: modification of work by Scott Bauer, USDA-ARS; credit b: modification of work by Stephen Ausmus, USDA-ARS; credit c: modification of work by David Marshall, USDA-ARS; credit d: modification of work by Joseph Smilanick, USDA-ARS)

Aflatoxins are toxic, carcinogenic compounds released by fungi of the genus *Aspergillus*. Periodically, harvests of nuts and grains are tainted by aflatoxins, leading to massive recall of produce. This sometimes ruins producers and causes food shortages in developing countries.

Animal and Human Parasites and Pathogens

Fungi can affect animals, including humans, in several ways. A **mycosis** is a fungal disease that results from infection and direct damage. Fungi attack animals directly by colonizing and destroying tissues. **Mycotoxicosis** is the

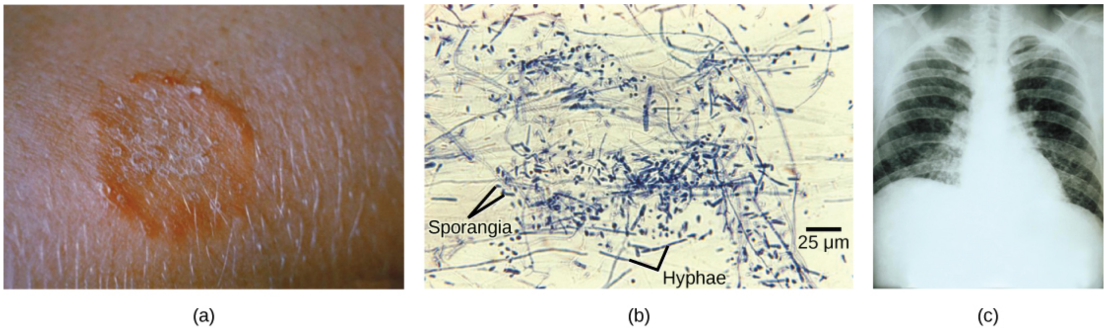
poisoning of humans (and other animals) by foods contaminated by fungal toxins (mycotoxins). **Mycetismus** describes the ingestion of preformed toxins in poisonous mushrooms. In addition, individuals who display hypersensitivity to molds and spores develop strong and dangerous allergic reactions. Fungal infections are generally very difficult to treat because, unlike bacteria, fungi are eukaryotes. Antibiotics only target prokaryotic cells, whereas compounds that kill fungi also harm the eukaryotic animal host.

Many fungal infections are superficial; that is, they occur on the animal's skin. Termed cutaneous ("skin") mycoses, they can have devastating effects. For example, the decline of the world's frog population in recent years may be caused by the chytrid fungus *Batrachochytrium dendrobatidis,* which infects the skin of frogs and presumably interferes with gaseous exchange. Similarly, more than a million bats in the United States have been killed by white-nose syndrome, which appears as a white ring around the mouth of the bat. It is caused by the cold-loving fungus *Pseudogymnoascus destructans*, which disseminates its deadly spores in caves where bats hibernate. Mycologists are researching the transmission, mechanism, and control of *P. destructans* to stop its spread.

Fungi that cause the superficial mycoses of the epidermis, hair, and nails rarely spread to the underlying tissue ([link]). These fungi are often misnamed "dermatophytes", from the Greek words *dermis* meaning skin and *phyte* meaning plant, although they are not plants. Dermatophytes are also called "ringworms" because of the red ring they cause on skin. They secrete extracellular enzymes that break down keratin (a protein found in hair, skin, and nails), causing conditions such as athlete's foot and jock itch. These conditions are usually treated with over-the-counter topical creams and powders, and are easily cleared. More persistent superficial mycoses may require prescription oral medications.

(a) Ringworm presents as a red ring on skin; (b) *Trichophyton violaceum*, shown in this bright field light micrograph, causes superficial mycoses on the scalp; (c) *Histoplasma capsulatum* is an

ascomycete that infects airways and causes symptoms similar to influenza. (credit a: modification of work by Dr. Lucille K. Georg, CDC; credit b: modification of work by Dr. Lucille K. Georg, CDC; credit c: modification of work by M. Renz, CDC; scale-bar data from Matt Russell)



Systemic mycoses spread to internal organs, most commonly entering the body through the respiratory system. For example, coccidioidomycosis (valley fever) is commonly found in the southwestern United States, where the fungus resides in the dust. Once inhaled, the spores develop in the lungs and cause symptoms similar to those of tuberculosis. Histoplasmosis is caused by the dimorphic fungus *Histoplasma capsulatum.* It also causes pulmonary infections, and in rarer cases, swelling of the membranes of the brain and spinal cord. Treatment of these and many other fungal diseases requires the use of antifungal medications that have serious side effects.

Opportunistic mycoses are fungal infections that are either common in all environments, or part of the normal biota. They mainly affect individuals who have a compromised immune system. Patients in the late stages of AIDS suffer from opportunistic mycoses that can be life threatening. The yeast *Candida* sp., a common member of the natural biota, can grow unchecked and infect the vagina or mouth (oral thrush) if the pH of the surrounding environment, the person's immune defenses, or the normal population of bacteria are altered.

Mycetismus can occur when poisonous mushrooms are eaten. It causes a number of human fatalities during mushroom-picking season. Many edible fruiting bodies of fungi resemble highly poisonous relatives, and amateur mushroom hunters are cautioned to carefully inspect their harvest and avoid

eating mushrooms of doubtful origin. The adage “there are bold mushroom pickers and old mushroom pickers, but are there no old, bold mushroom pickers” is unfortunately true.

Note:

Scientific Method Connection

Dutch Elm Disease

Question: Do trees resistant to Dutch elm disease secrete antifungal compounds?

Hypothesis: Construct a hypothesis that addresses this question.

Background: Dutch elm disease is a fungal infestation that affects many species of elm (*Ulmus*) in North America. The fungus infects the vascular system of the tree, which blocks water flow within the plant and mimics drought stress. Accidently introduced to the United States in the early 1930s, it decimated shade trees across the continent. It is caused by the fungus *Ophiostoma ulmi*. The elm bark beetle acts as a vector and transmits the disease from tree to tree. Many European and Asiatic elms are less susceptible to the disease than are American elms.

Test the hypothesis: A researcher testing this hypothesis might do the following. Inoculate several Petri plates containing a medium that supports the growth of fungi with fragments of *Ophiostoma* mycelium. Cut (with a metal punch) several disks from the vascular tissue of susceptible varieties of American elms and resistant European and Asiatic elms. Include control Petri plates inoculated with mycelia without plant tissue to verify that the medium and incubation conditions do not interfere with fungal growth. As a positive control, add paper disks impregnated with a known fungicide to Petri plates inoculated with the mycelium.

Incubate the plates for a set number of days to allow fungal growth and spreading of the mycelium over the surface of the plate. Record the diameter of the zone of clearing, if any, around the tissue samples and the fungicide control disk.

Record your observations in the following table.

Results of Antifungal Testing of Vascular Tissue from Different Species of Elm	
Disk	**Zone of Inhibition (mm)**
Distilled Water	
Fungicide	
Tissue from Susceptible Elm #1	
Tissue from Susceptible Elm #2	
Tissue from Resistant Elm #1	
Tissue from Resistant Elm #2	

Analyze the data and report the results. Compare the effect of distilled water to the fungicide. These are negative and positive controls that validate the experimental set up. The fungicide should be surrounded by a clear zone where the fungus growth was inhibited. Is there a difference among different species of elm?
Draw a conclusion: Was there antifungal activity as expected from the fungicide? Did the results support the hypothesis? If not, how can this be explained? There are several possible explanations for resistance to a pathogen. Active deterrence of infection is only one of them.

Section Summary

Fungi establish parasitic relationships with plants and animals. Fungal diseases can decimate crops and spoil food during storage. Compounds produced by fungi can be toxic to humans and other animals. Mycoses are infections caused by fungi. Superficial mycoses affect the skin, whereas

systemic mycoses spread through the body. Fungal infections are difficult to cure.

Review Questions

Exercise:

Problem:

A fungus that climbs up a tree reaching higher elevation to release its spores in the wind and does not receive any nutrients from the tree or contribute to the tree's welfare is described as a _________.

a. commensal
b. mutualist
c. parasite
d. pathogen

Solution:

A

Exercise:

Problem:

A fungal infection that affects nails and skin is classified as _________.

a. systemic mycosis
b. mycetismus
c. superficial mycosis
d. mycotoxicosis

Solution:

C

Free Response

Exercise:

Problem:

Why can superficial mycoses in humans lead to bacterial infections?

Solution:

Dermatophytes that colonize skin break down the keratinized layer of dead cells that protects tissues from bacterial invasion. Once the integrity of the skin is breached, bacteria can enter the deeper layers of tissues and cause infections.

Glossary

commensalism
: symbiotic relationship in which one member benefits while the other member is not affected

mycetismus
: ingestion of toxins in poisonous mushrooms

mycosis
: fungal infection

mycotoxicosis
: poisoning by a fungal toxin released in food

parasitism
: symbiotic relationship in which one member of the association benefits at the expense of the other

Importance of Fungi in Human Life

By the end of this section, you will be able to:

- Describe the importance of fungi to the balance of the environment
- Summarize the role of fungi in food and beverage preparation
- Describe the importance of fungi in the chemical and pharmaceutical industries
- Discuss the role of fungi as model organisms

Although we often think of fungi as organisms that cause disease and rot food, fungi are important to human life on many levels. As we have seen, they influence the well-being of human populations on a large scale because they are part of the nutrient cycle in ecosystems. They have other ecosystem roles as well. As animal pathogens, fungi help to control the population of damaging pests. These fungi are very specific to the insects they attack, and do not infect animals or plants. Fungi are currently under investigation as potential microbial insecticides, with several already on the market. For example, the fungus *Beauveria bassiana* is a pesticide being tested as a possible biological control agent for the recent spread of emerald ash borer. It has been released in Michigan, Illinois, Indiana, Ohio, West Virginia and Maryland ([link]).

The emerald ash borer is an insect that attacks ash trees. It is in turn parasitized by a pathogenic fungus

that holds promise as a biological insecticide. The parasitic fungus appears as white fuzz on the body of the insect. (credit: Houping Liu, USDA Agricultural Research Service)

The mycorrhizal relationship between fungi and plant roots is essential for the productivity of farm land. Without the fungal partner in root systems, 80–90 percent of trees and grasses would not survive. Mycorrhizal fungal inoculants are available as soil amendments from gardening supply stores and are promoted by supporters of organic agriculture.

We also eat some types of fungi. Mushrooms figure prominently in the human diet. Morels, shiitake mushrooms, chanterelles, and truffles are considered delicacies ([link]). The humble meadow mushroom, *Agaricus campestris*, appears in many dishes. Molds of the genus *Penicillium* ripen many cheeses. They originate in the natural environment such as the caves of Roquefort, France, where wheels of sheep milk cheese are stacked in order to capture the molds responsible for the blue veins and pungent taste of the cheese.

The morel mushroom is an ascomycete much appreciated for its delicate taste. (credit: Jason Hollinger)

Fermentation—of grains to produce beer, and of fruits to produce wine—is an ancient art that humans in most cultures have practiced for millennia. Wild yeasts are acquired from the environment and used to ferment sugars into CO_2 and ethyl alcohol under anaerobic conditions. It is now possible to purchase isolated strains of wild yeasts from different wine-making regions. Louis Pasteur was instrumental in developing a reliable strain of brewer's yeast, *Saccharomyces cerevisiae*, for the French brewing industry in the late 1850s. This was one of the first examples of biotechnology patenting.

Many secondary metabolites of fungi are of great commercial importance. Antibiotics are naturally produced by fungi to kill or inhibit the growth of bacteria, limiting their competition in the natural environment. Important antibiotics, such as penicillin and the cephalosporins, are isolated from fungi. Valuable drugs isolated from fungi include the immunosuppressant drug cyclosporine (which reduces the risk of rejection after organ transplant), the precursors of steroid hormones, and ergot alkaloids used to stop bleeding. Psilocybin is a compound found in fungi such as *Psilocybe semilanceata* and *Gymnopilus junonius*, which have been used for their hallucinogenic properties by various cultures for thousands of years.

As simple eukaryotic organisms, fungi are important model research organisms. Many advances in modern genetics were achieved by the use of the red bread mold *Neurospora crassa*. Additionally, many important genes originally discovered in *S. cerevisiae* served as a starting point in discovering analogous human genes. As a eukaryotic organism, the yeast cell produces and modifies proteins in a manner similar to human cells, as opposed to the bacterium *Escherichia coli*, which lacks the internal membrane structures and enzymes to tag proteins for export. This makes yeast a much better organism for use in recombinant DNA technology

experiments. Like bacteria, yeasts grow easily in culture, have a short generation time, and are amenable to genetic modification.

Section Summary

Fungi are important to everyday human life. Fungi are important decomposers in most ecosystems. Mycorrhizal fungi are essential for the growth of most plants. Fungi, as food, play a role in human nutrition in the form of mushrooms, and also as agents of fermentation in the production of bread, cheeses, alcoholic beverages, and numerous other food preparations. Secondary metabolites of fungi are used as medicines, such as antibiotics and anticoagulants. Fungi are model organisms for the study of eukaryotic genetics and metabolism.

Review Questions

Exercise:

Problem:

Yeast is a facultative anaerobe. This means that alcohol fermentation takes place only if:

a. the temperature is close to 37°C
b. the atmosphere does not contain oxygen
c. sugar is provided to the cells
d. light is provided to the cells

Solution:

B

Exercise:

Problem:

The advantage of yeast cells over bacterial cells to express human proteins is that:

a. yeast cells grow faster
b. yeast cells are easier to manipulate genetically
c. yeast cells are eukaryotic and modify proteins similarly to human cells
d. yeast cells are easily lysed to purify the proteins

Solution:

C

Free Response

Exercise:

Problem:

Historically, artisanal breads were produced by capturing wild yeasts from the air. Prior to the development of modern yeast strains, the production of artisanal breads was long and laborious because many batches of dough ended up being discarded. Can you explain this fact?

Solution:

The dough is often contaminated by toxic spores that float in the air. It was one of Louis Pasteur’s achievements to purify reliable strains of baker’s yeast to produce bread consistently.

Early Plant Life

By the end of this section, you will be able to:

- Discuss the challenges to plant life on land
- Describe the adaptations that allowed plants to colonize the land
- Describe the timeline of plant evolution and the impact of land plants on other living things

The kingdom Plantae constitutes large and varied groups of organisms. There are more than 300,000 species of catalogued plants. Of these, more than 260,000 are seed plants. Mosses, ferns, conifers, and flowering plants are all members of the plant kingdom. Most biologists also consider green algae to be plants, although others exclude all algae from the plant kingdom. The reason for this disagreement stems from the fact that only green algae, the **Charophytes**, share common characteristics with land plants (such as using chlorophyll *a* and *b* plus carotene in the same proportion as plants). These characteristics are absent in other types of algae.

Note:

Evolution Connection

Algae and Evolutionary Paths to Photosynthesis

Some scientists consider all algae to be plants, while others assert that only the Charophytes belong in the kingdom Plantae. These divergent opinions are related to the different evolutionary paths to photosynthesis selected for in different types of algae. While all algae are photosynthetic—that is, they contain some form of a chloroplast—they didn’t all become photosynthetic via the same path.

The ancestors to the green algae became photosynthetic by endosymbiosing a green, photosynthetic bacterium about 1.65 billion years ago. That algal line evolved into the Charophytes, and eventually into the modern mosses, ferns, gymnosperms, and angiosperms. Their evolutionary trajectory was relatively straight and monophyletic. In contrast, the other algae—red, brown, golden, stramenopiles, and so on—all became photosynthetic by secondary, or even tertiary, endosymbiotic events; that is, they endosymbiosed cells that had already endosymbiosed a

cyanobacterium. These latecomers to photosynthesis are parallels to the Charophytes in terms of autotrophy, but they did not expand to the same extent as the Charophytes, nor did they colonize the land.
The different views on whether all algae are Plantae arise from how these evolutionary paths are viewed. Scientists who solely track evolutionary straight lines (that is, monophyly), consider only the Charophytes as plants. To biologists who cast a broad net over living things that share a common characteristic (in this case, photosynthetic eukaryotes), all algae are plants.

Note:
Link to Learning

Go to this interactive website to get a more in-depth view of the Charophytes.

Plant Adaptations to Life on Land

As organisms adapted to life on land, they had to contend with several challenges in the terrestrial environment. Water has been described as "the stuff of life." The cell's interior is a watery soup: in this medium, most small molecules dissolve and diffuse, and the majority of the chemical reactions of metabolism take place. Desiccation, or drying out, is a constant danger for an organism exposed to air. Even when parts of a plant are close to a source of water, the aerial structures are likely to dry out. Water also provides buoyancy to organisms. On land, plants need to develop structural support in a medium that does not give the same lift as water. The organism is also subject to bombardment by mutagenic radiation, because air does not

filter out ultraviolet rays of sunlight. Additionally, the male gametes must reach the female gametes using new strategies, because swimming is no longer possible. Therefore, both gametes and zygotes must be protected from desiccation. The successful land plants developed strategies to deal with all of these challenges. Not all adaptations appeared at once. Some species never moved very far from the aquatic environment, whereas others went on to conquer the driest environments on Earth.

To balance these survival challenges, life on land offers several advantages. First, sunlight is abundant. Water acts as a filter, altering the spectral quality of light absorbed by the photosynthetic pigment chlorophyll. Second, carbon dioxide is more readily available in air than in water, since it diffuses faster in air. Third, land plants evolved before land animals; therefore, until dry land was colonized by animals, no predators threatened plant life. This situation changed as animals emerged from the water and fed on the abundant sources of nutrients in the established flora. In turn, plants developed strategies to deter predation: from spines and thorns to toxic chemicals.

Early land plants, like the early land animals, did not live very far from an abundant source of water and developed survival strategies to combat dryness. One of these strategies is called tolerance. Many mosses, for example, can dry out to a brown and brittle mat, but as soon as rain or a flood makes water available, mosses will absorb it and are restored to their healthy green appearance. Another strategy is to colonize environments with high humidity, where droughts are uncommon. Ferns, which are considered an early lineage of plants, thrive in damp and cool places such as the understory of temperate forests. Later, plants moved away from moist or aquatic environments using resistance to desiccation, rather than tolerance. These plants, like cacti, minimize the loss of water to such an extent they can survive in extremely dry environments.

Additional Land Plant Adaptations

As plants adapted to dry land and became independent from the constant presence of water in damp habitats, new organs and structures made their appearance. Early land plants did not grow more than a few inches off the

ground, competing for light on these low mats. By developing a shoot and growing taller, individual plants captured more light. Because air offers substantially less support than water, land plants incorporated more rigid molecules in their stems (and later, tree trunks). In small plants such as single-celled algae, simple diffusion suffices to distribute water and nutrients throughout the organism. However, for plants to evolve larger forms, the evolution of vascular tissue for the distribution of water and solutes was a prerequisite. The vascular system contains xylem and phloem tissues. Xylem conducts water and minerals absorbed from the soil up to the shoot, while phloem transports food derived from photosynthesis throughout the entire plant. A root system evolved to take up water and minerals from the soil, and to anchor the increasingly taller shoot in the soil.

In land plants, a waxy, waterproof cover called a cuticle protects the leaves and stems from desiccation. However, the cuticle also prevents intake of carbon dioxide needed for the synthesis of carbohydrates through photosynthesis. To overcome this, stomata or pores that open and close to regulate traffic of gases and water vapor appeared in plants as they moved away from moist environments into drier habitats.

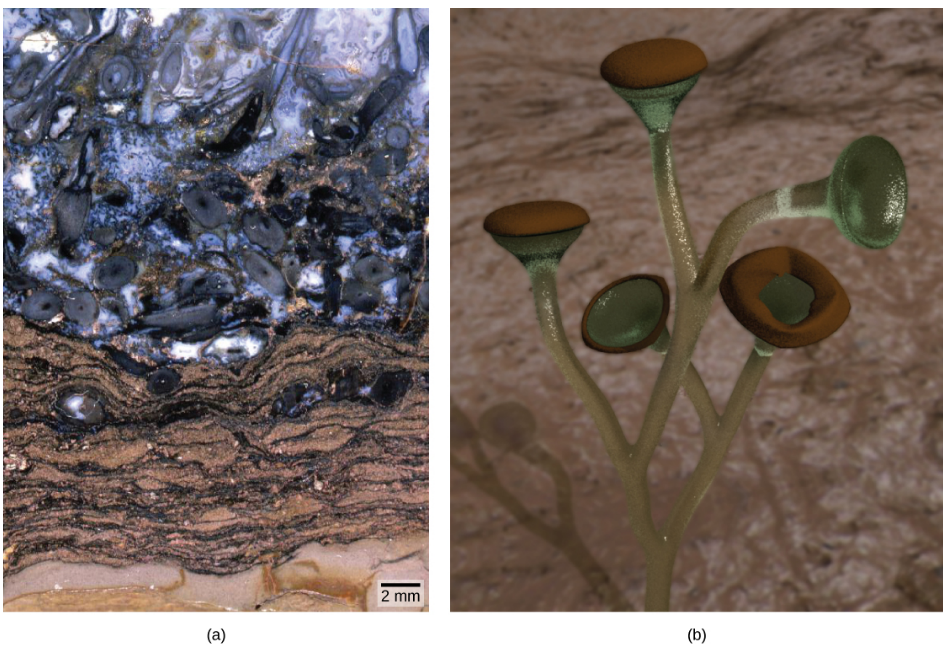
Water filters ultraviolet-B (UVB) light, which is harmful to all organisms, especially those that must absorb light to survive. This filtering does not occur for land plants. This presented an additional challenge to land colonization, which was met by the evolution of biosynthetic pathways for the synthesis of protective flavonoids and other compounds: pigments that absorb UV wavelengths of light and protect the aerial parts of plants from photodynamic damage.

Plants cannot avoid being eaten by animals. Instead, they synthesize a large range of poisonous secondary metabolites: complex organic molecules such as alkaloids, whose noxious smells and unpleasant taste deter animals. These toxic compounds can also cause severe diseases and even death, thus discouraging predation. Humans have used many of these compounds for centuries as drugs, medications, or spices. In contrast, as plants co-evolved with animals, the development of sweet and nutritious metabolites lured animals into providing valuable assistance in dispersing pollen grains, fruit,

or seeds. Plants have been enlisting animals to be their helpers in this way for hundreds of millions of years.

Evolution of Land Plants

No discussion of the evolution of plants on land can be undertaken without a brief review of the timeline of the geological eras. The early era, known as the Paleozoic, is divided into six periods. It starts with the Cambrian period, followed by the Ordovician, Silurian, Devonian, Carboniferous, and Permian. The major event to mark the Ordovician, more than 500 million years ago, was the colonization of land by the ancestors of modern land plants. Fossilized cells, cuticles, and spores of early land plants have been dated as far back as the Ordovician period in the early Paleozoic era. The oldest-known vascular plants have been identified in deposits from the Devonian. One of the richest sources of information is the Rhynie chert, a sedimentary rock deposit found in Rhynie, Scotland ([link]), where embedded fossils of some of the earliest vascular plants have been identified.



This Rhynie chert contains fossilized material from

vascular plants. The area inside the circle contains bulbous underground stems called corms, and root-like structures called rhizoids. (credit b: modification of work by Peter Coxhead based on original image by “Smith609”/Wikimedia Commons; scale-bar data from Matt Russell)

Paleobotanists distinguish between **extinct** species, as fossils, and **extant** species, which are still living. The extinct vascular plants, classified as zosterophylls and trimerophytes, most probably lacked true leaves and roots and formed low vegetation mats similar in size to modern-day mosses, although some trimetophytes could reach one meter in height. The later genus *Cooksonia*, which flourished during the Silurian, has been extensively studied from well-preserved examples. Imprints of *Cooksonia* show slender branching stems ending in what appear to be sporangia. From the recovered specimens, it is not possible to establish for certain whether *Cooksonia* possessed vascular tissues. Fossils indicate that by the end of the Devonian period, ferns, horsetails, and seed plants populated the landscape, giving rising to trees and forests. This luxuriant vegetation helped enrich the atmosphere in oxygen, making it easier for air-breathing animals to colonize dry land. Plants also established early symbiotic relationships with fungi, creating mycorrhizae: a relationship in which the fungal network of filaments increases the efficiency of the plant root system, and the plants provide the fungi with byproducts of photosynthesis.

Note:
Career Connection
Paleobotanist
How organisms acquired traits that allow them to colonize new environments—and how the contemporary ecosystem is shaped—are fundamental questions of evolution. Paleobotany (the study of extinct plants) addresses these questions through the analysis of fossilized specimens retrieved from field studies, reconstituting the morphology of

organisms that disappeared long ago. Paleobotanists trace the evolution of plants by following the modifications in plant morphology: shedding light on the connection between existing plants by identifying common ancestors that display the same traits. This field seeks to find transitional species that bridge gaps in the path to the development of modern organisms. Fossils are formed when organisms are trapped in sediments or environments where their shapes are preserved. Paleobotanists collect fossil specimens in the field and place them in the context of the geological sediments and other fossilized organisms surrounding them. The activity requires great care to preserve the integrity of the delicate fossils and the layers of rock in which they are found.

One of the most exciting recent developments in paleobotany is the use of analytical chemistry and molecular biology to study fossils. Preservation of molecular structures requires an environment free of oxygen, since oxidation and degradation of material through the activity of microorganisms depend on its presence. One example of the use of analytical chemistry and molecular biology is the identification of oleanane, a compound that deters pests. Up to this point, oleanane appeared to be unique to flowering plants; however, it has now been recovered from sediments dating from the Permian, much earlier than the current dates given for the appearance of the first flowering plants. Paleobotanists can also study fossil DNA, which can yield a large amount of information, by analyzing and comparing the DNA sequences of extinct plants with those of living and related organisms. Through this analysis, evolutionary relationships can be built for plant lineages.

Some paleobotanists are skeptical of the conclusions drawn from the analysis of molecular fossils. For example, the chemical materials of interest degrade rapidly when exposed to air during their initial isolation, as well as in further manipulations. There is always a high risk of contaminating the specimens with extraneous material, mostly from microorganisms. Nevertheless, as technology is refined, the analysis of DNA from fossilized plants will provide invaluable information on the evolution of plants and their adaptation to an ever-changing environment.

The Major Divisions of Land Plants

The green algae and land plants are grouped together into a subphylum called the Streptophytina, and thus are called Streptophytes. In a further division, land plants are classified into two major groups according to the absence or presence of vascular tissue, as detailed in [link]. Plants that lack vascular tissue, which is formed of specialized cells for the transport of water and nutrients, are referred to as **non-vascular plants**. Liverworts, mosses, and hornworts are seedless, non-vascular plants that likely appeared early in land plant evolution. Vascular plants developed a network of cells that conduct water and solutes. The first vascular plants appeared in the late Ordovician and were probably similar to lycophytes, which include club mosses (not to be confused with the mosses) and the pterophytes (ferns, horsetails, and whisk ferns). Lycophytes and pterophytes are referred to as seedless vascular plants, because they do not produce seeds. The seed plants, or spermatophytes, form the largest group of all existing plants, and hence dominate the landscape. Seed plants include gymnosperms, most notably conifers (Gymnosperms), which produce "naked seeds," and the most successful of all plants, the flowering plants (Angiosperms). Angiosperms protect their seeds inside chambers at the center of a flower; the walls of the chamber later develop into a fruit.

Note:
Art Connection

<table>
<tr><th colspan="8">STREPTOPHYTES: THE GREEN PLANTS</th></tr>
<tr><td rowspan="7">Charophytes</td><td colspan="7">Embryophytes: The Land Plants</td></tr>
<tr><td colspan="3">Non Vascular</td><td colspan="4">Vascular</td></tr>
<tr><td colspan="3" rowspan="2">Seedless Plants
Bryophytes</td><td colspan="2">Seedless Plants</td><td colspan="2" rowspan="2">Seed Plants
Spermatophytes</td></tr>
<tr><td>Lycophytes</td><td>Pterophytes</td></tr>
<tr><td rowspan="3">Liver-worts</td><td rowspan="3">Horn-worts</td><td rowspan="3">Mosses</td><td>Club Mosses</td><td>Whisk Ferns</td><td rowspan="3">Gymno-sperms</td><td rowspan="3">Angio-sperms</td></tr>
<tr><td>Quillworts</td><td>Horsetails</td></tr>
<tr><td>Spike Mosses</td><td>Ferns</td></tr>
</table>

This table shows the major divisions of green plants.

Which of the following statements about plant divisions is false?

a. Lycophytes and pterophytes are seedless vascular plants.
b. All vascular plants produce seeds.
c. All nonvascular embryophytes are bryophytes.
d. Seed plants include angiosperms and gymnosperms.

Section Summary

Land plants acquired traits that made it possible to colonize land and survive out of the water. All land plants share the following characteristics: alternation of generations, with the haploid plant called a gametophyte, and the diploid plant called a sporophyte; protection of the embryo, formation of haploid spores in a sporangium, formation of gametes in a gametangium, and an apical meristem. Vascular tissues, roots, leaves, cuticle cover, and a tough outer layer that protects the spores contributed to the adaptation of plants to dry land. Land plants appeared about 500 million years ago in the Ordovician period.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about plant divisions is false?

a. Lycophytes and pterophytes are seedless vascular plants.
b. All vascular plants produce seeds.
c. All nonvascular embryophytes are bryophytes.
d. Seed plants include angiosperms and gymnosperms.

Solution:

[link] B.

Review Questions

Exercise:

Problem:

The land plants are probably descendants of which of these groups?

a. green algae
b. red algae
c. brown algae
d. angiosperms

Solution:

A

Exercise:

Problem: Alternation of generations means that plants produce:

a. only haploid multicellular organisms
b. only diploid multicellular organisms
c. only diploid multicellular organisms with single-celled haploid gametes
d. both haploid and diploid multicellular organisms

Solution:

D

Exercise:

Problem:

Which of the following traits of land plants allows them to grow in height?

a. alternation of generations
b. waxy cuticle
c. tracheids
d. sporopollenin

Solution:

C

Free Response

Exercise:

Problem:

Why did land plants lose some of the accessory pigments present in brown and red algae?

Solution:

Sunlight is not filtered by water or other algae on land; therefore, there is no need to collect light at additional wavelengths made available by other pigment coloration.

Exercise:

Problem: What is the difference between extant and extinct?

Solution:

Paleobotanists distinguish between extinct species, which no longer live, and extant species, which are still living.

Glossary

antheridium
: male gametangium

archegonium
: female gametangium

charophyte
: other term for green algae; considered the closest relative of land plants

diplontic
: diploid stage is the dominant stage

embryophyte
: other name for land plant; embryo is protected and nourished by the sporophyte

extant
: still-living species

extinct
: no longer existing species

gametangium
: structure on the gametophyte in which gametes are produced

haplodiplodontic
: haploid and diploid stages alternate

haplontic
: haploid stage is the dominant stage

heterosporous
: produces two types of spores

homosporous
: produces one type of spore

megaspore
: female spore

microspore
: male spore

non-vascular plant
: plant that lacks vascular tissue, which is formed of specialized cells for the transport of water and nutrients

seedless vascular plant
: plant that does not produce seeds

sporocyte
: diploid cell that produces spores by meiosis

sporopollenin
: tough polymer surrounding the spore

vascular plant
: plant containing a network of cells that conducts water and solutes through the organism

Bryophytes

By the end of this section, you will be able to:

- Identify the main characteristics of bryophytes
- Describe the distinguishing traits of liverworts, hornworts, and mosses
- Chart the development of land adaptations in the bryophytes
- Describe the events in the bryophyte lifecycle

Bryophytes are the group of plants that are the closest extant relative of early terrestrial plants. The first bryophytes (liverworts) most likely appeared in the Ordovician period, about 450 million years ago. Because of the lack of lignin and other resistant structures, the likelihood of bryophytes forming fossils is rather small. Some spores protected by sporopollenin have survived and are attributed to early bryophytes. By the Silurian period, however, vascular plants had spread through the continents. This compelling fact is used as evidence that non-vascular plants must have preceded the Silurian period.

More than 25,000 species of bryophytes thrive in mostly damp habitats, although some live in deserts. They constitute the major flora of inhospitable environments like the tundra, where their small size and tolerance to desiccation offer distinct advantages. They generally lack lignin and do not have actual tracheids (xylem cells specialized for water conduction). Rather, water and nutrients circulate inside specialized conducting cells. Although the term non-tracheophyte is more accurate, bryophytes are commonly called nonvascular plants.

In a bryophyte, all the conspicuous vegetative organs—including the photosynthetic leaf-like structures, the thallus, stem, and the rhizoid that anchors the plant to its substrate—belong to the haploid organism or gametophyte. The sporophyte is barely noticeable. The gametes formed by bryophytes swim with a flagellum, as do gametes in a few of the tracheophytes. The sporangium—the multicellular sexual reproductive structure—is present in bryophytes and absent in the majority of algae. The bryophyte embryo also remains attached to the parent plant, which protects and nourishes it. This is a characteristic of land plants.

The bryophytes are divided into three phyla: the liverworts or Hepaticophyta, the hornworts or Anthocerotophyta, and the mosses or true Bryophyta.

Liverworts

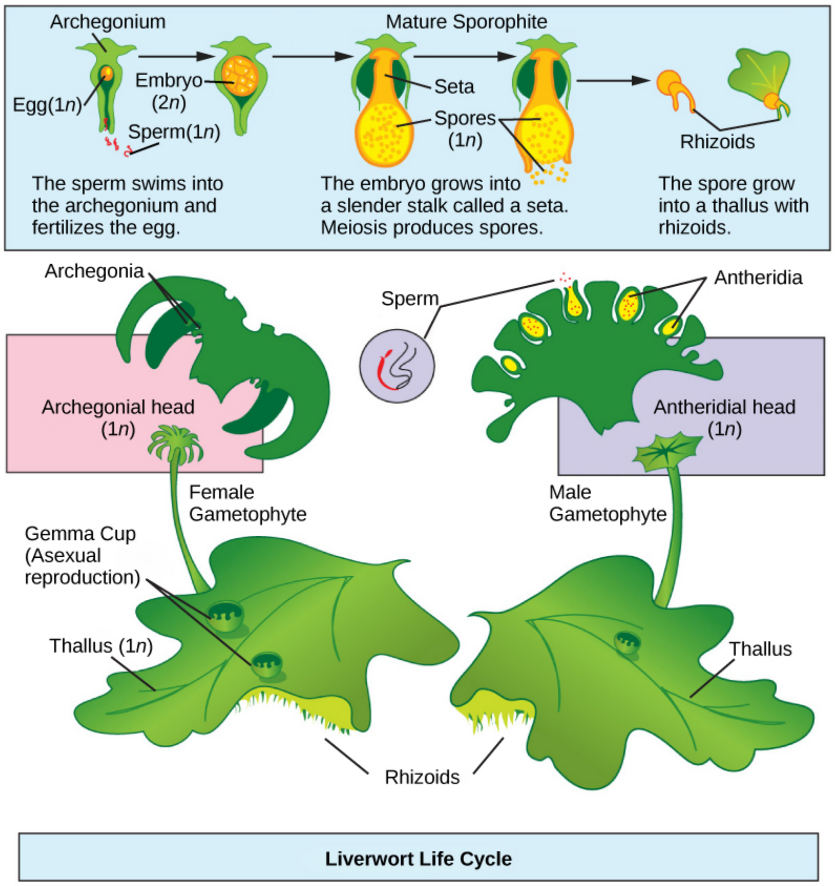
Liverworts (Hepaticophyta) are viewed as the plants most closely related to the ancestor that moved to land. Liverworts have colonized every terrestrial habitat on Earth and diversified to more than 7000 existing species ([link]). Some gametophytes form lobate green structures, as seen in [link]. The shape is similar to the lobes of the liver, and hence provides the origin of the name given to the phylum.

This 1904 drawing shows the variety of forms of Hepaticophyta.

A liverwort, *Lunularia cruciata*, displays its lobate, flat thallus. The organism in the photograph is in the gametophyte stage.

Openings that allow the movement of gases may be observed in liverworts. However, these are not stomata, because they do not actively open and close. The plant takes up water over its entire surface and has no cuticle to prevent desiccation. [link] represents the lifecycle of a liverwort. The cycle starts with the release of haploid spores from the sporangium that developed on the sporophyte. Spores disseminated by wind or water germinate into flattened thalli attached to the substrate by thin, single-celled filaments. Male and female gametangia develop on separate, individual plants. Once released, male gametes swim with the aid of their flagella to the female gametangium (the archegonium), and fertilization ensues. The zygote grows into a small sporophyte still attached to the parent gametophyte. It will give rise, by meiosis, to the next generation of spores. Liverwort plants can also reproduce asexually, by the breaking of branches or the spreading of leaf fragments called gemmae. In this latter type of reproduction, the **gemmae** —small, intact, complete pieces of plant that are produced in a cup on the

surface of the thallus (shown in [link])—are splashed out of the cup by raindrops. The gemmae then land nearby and develop into gametophytes.



The life cycle of a typical liverwort is shown. (credit: modification of work by Mariana Ruiz Villareal)

Hornworts

The **hornworts** (*Anthocerotophyta*) belong to the broad bryophyte group. They have colonized a variety of habitats on land, although they are never far from a source of moisture. The short, blue-green gametophyte is the dominant phase of the lifecycle of a hornwort. The narrow, pipe-like

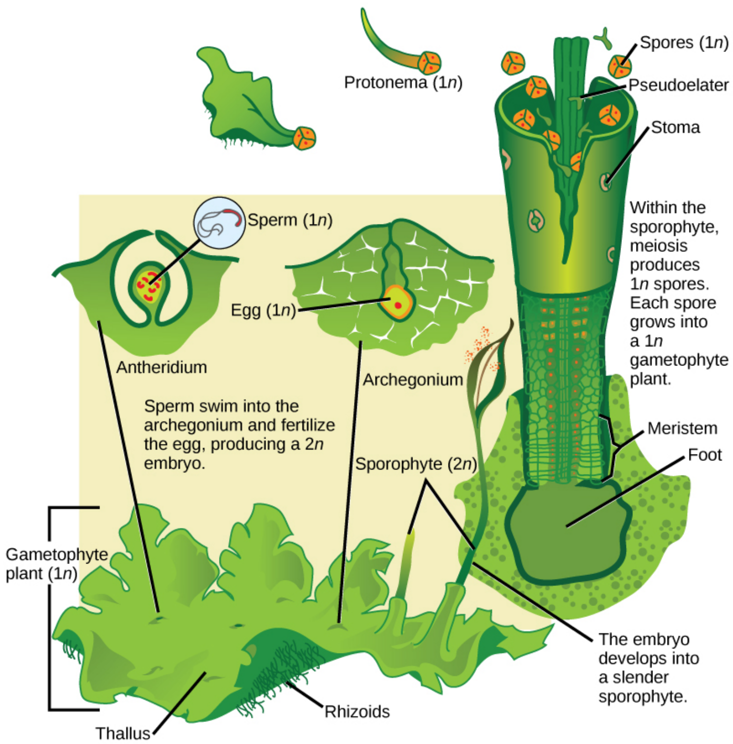
sporophyte is the defining characteristic of the group. The sporophytes emerge from the parent gametophyte and continue to grow throughout the life of the plant ([link]).

Hornworts grow a tall and slender sporophyte. (credit: modification of work by Jason Hollinger)

Stomata appear in the hornworts and are abundant on the sporophyte. Photosynthetic cells in the thallus contain a single chloroplast. Meristem cells at the base of the plant keep dividing and adding to its height. Many hornworts establish symbiotic relationships with cyanobacteria that fix nitrogen from the environment.

The lifecycle of hornworts ([link]) follows the general pattern of alternation of generations. The gametophytes grow as flat thalli on the soil with embedded gametangia. Flagellated sperm swim to the archegonia and fertilize eggs. The zygote develops into a long and slender sporophyte that eventually splits open, releasing spores. Thin cells called pseudoelaters surround the spores and help propel them further in the environment. Unlike the elaters observed in horsetails, the hornwort pseudoelaters are

single-celled structures. The haploid spores germinate and give rise to the next generation of gametophyte.



The alternation of generation in hornworts is shown. (credit: modification of work by “Smith609”/Wikimedia Commons based on original work by Mariana Ruiz Villareal)

Mosses

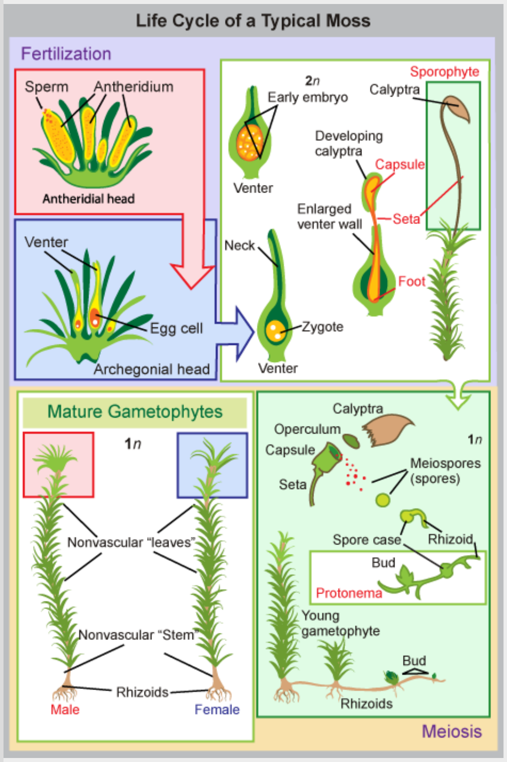
More than 10,000 species of **mosses** have been catalogued. Their habitats vary from the tundra, where they are the main vegetation, to the understory of tropical forests. In the tundra, the mosses’ shallow rhizoids allow them to fasten to a substrate without penetrating the frozen soil. Mosses slow down

erosion, store moisture and soil nutrients, and provide shelter for small animals as well as food for larger herbivores, such as the musk ox. Mosses are very sensitive to air pollution and are used to monitor air quality. They are also sensitive to copper salts, so these salts are a common ingredient of compounds marketed to eliminate mosses from lawns.

Mosses form diminutive gametophytes, which are the dominant phase of the lifecycle. Green, flat structures—resembling true leaves, but lacking vascular tissue—are attached in a spiral to a central stalk. The plants absorb water and nutrients directly through these leaf-like structures. Some mosses have small branches. Some primitive traits of green algae, such as flagellated sperm, are still present in mosses that are dependent on water for reproduction. Other features of mosses are clearly adaptations to dry land. For example, stomata are present on the stems of the sporophyte, and a primitive vascular system runs up the sporophyte’s stalk. Additionally, mosses are anchored to the substrate—whether it is soil, rock, or roof tiles—by multicellular **rhizoids**. These structures are precursors of roots. They originate from the base of the gametophyte, but are not the major route for the absorption of water and minerals. The lack of a true root system explains why it is so easy to rip moss mats from a tree trunk. The moss lifecycle follows the pattern of alternation of generations as shown in [link]. The most familiar structure is the haploid gametophyte, which germinates from a haploid spore and forms first a **protonema**—usually, a tangle of single-celled filaments that hug the ground. Cells akin to an apical meristem actively divide and give rise to a gametophore, consisting of a photosynthetic stem and foliage-like structures. Rhizoids form at the base of the gametophore. Gametangia of both sexes develop on separate gametophores. The male organ (the antheridium) produces many sperm, whereas the archegonium (the female organ) forms a single egg. At fertilization, the sperm swims down the neck to the venter and unites with the egg inside the archegonium. The zygote, protected by the archegonium, divides and grows into a sporophyte, still attached by its foot to the gametophyte.

Note:

Art Connection



This illustration shows the life cycle of mosses. (credit: modification of work by Mariana Ruiz Villareal)

Which of the following statements about the moss life cycle is false?

a. The mature gametophyte is haploid.
b. The sporophyte produces haploid spores.
c. The calyptra buds to form a mature gametophyte.
d. The zygote is housed in the venter.

The slender **seta** (plural, setae), as seen in [link], contains tubular cells that transfer nutrients from the base of the sporophyte (the foot) to the

sporangium or **capsule**.

This photograph shows the long slender stems, called setae, connected to capsules of the moss *Thamnobryum alopecurum*. (credit: modification of work by Hermann Schachner)

A structure called a **peristome** increases the spread of spores after the tip of the capsule falls off at dispersal. The concentric tissue around the mouth of the capsule is made of triangular, close-fitting units, a little like “teeth”; these open and close depending on moisture levels, and periodically release spores.

Section Summary

Seedless nonvascular plants are small, having the gametophyte as the dominant stage of the lifecycle. Without a vascular system and roots, they absorb water and nutrients on all their exposed surfaces. Collectively known as bryophytes, the three main groups include the liverworts, the hornworts, and the mosses. Liverworts are the most primitive plants and are

closely related to the first land plants. Hornworts developed stomata and possess a single chloroplast per cell. Mosses have simple conductive cells and are attached to the substrate by rhizoids. They colonize harsh habitats and can regain moisture after drying out. The moss sporangium is a complex structure that allows release of spores away from the parent plant.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the moss life cycle is false?

a. The mature gametophyte is haploid.
b. The sporophyte produces haploid spores.
c. The rhizoid buds to form a mature gametophyte.
d. The zygote is housed in the venter.

Solution:

[link] C.

Review Questions

Exercise:

Problem:

Which of the following structures is not found in bryophytes?

a. a cellulose cell wall
b. chloroplast
c. sporangium
d. root

Solution:

D

Exercise:

Problem: Stomata appear in which group of plants?

a. Charales
b. liverworts
c. hornworts
d. mosses

Solution:

C

Exercise:

Problem: The chromosome complement in a moss protonema is:

a. 1*n*
b. 2*n*
c. 3*n*
d. varies with the size of the protonema

Solution:

A

Exercise:

Problem: Why do mosses grow well in the Arctic tundra?

a. They grow better at cold temperatures.
b. They do not require moisture.

c. They do not have true roots and can grow on hard surfaces.
d. There are no herbivores in the tundra.

Solution:

C

Free Response

Exercise:

Problem:

In areas where it rains often, mosses grow on roofs. How do mosses survive on roofs without soil?

Solution:

Mosses absorb water and nutrients carried by the rain and do not need soil because they do not derive much nutrition from the soil.

Exercise:

Problem: What are the three classes of bryophytes?

Solution:

The bryophytes are divided into three phyla: the liverworts or Hepaticophyta, the hornworts or Anthocerotophyta, and the mosses or true Bryophyta.

Glossary

capsule
: case of the sporangium in mosses

gemma
: (plural, gemmae) leaf fragment that spreads for asexual reproduction

hornworts
: group of non-vascular plants in which stomata appear

liverworts
: most primitive group of the non-vascular plants

mosses
: group of bryophytes in which a primitive conductive system appears

peristome
: tissue that surrounds the opening of the capsule and allows periodic release of spores

protonema
: tangle of single celled filaments that forms from the haploid spore

rhizoids
: thin filaments that anchor the plant to the substrate

seta
: stalk that supports the capsule in mosses

Seedless Vascular Plants
By the end of this section, you will be able to:

- Identify the new traits that first appear in tracheophytes
- Discuss the importance of adaptations to life on land
- Describe the classes of seedless tracheophytes
- Explain the role of seedless vascular plants in the ecosystem

Bryophytes may have been successful at the transition from an aquatic habitat to land, but they are still dependent on water for reproduction, and absorb moisture and nutrients through the gametophyte surface. The lack of roots for absorbing water and minerals from the soil, as well as a lack of reinforced conducting cells, limits bryophytes to small sizes. Although they may survive in reasonably dry conditions, they cannot reproduce and expand their habitat range in the absence of water. Vascular plants, on the other hand, can achieve enormous heights, thus competing successfully for light. Photosynthetic organs become leaves, and pipe-like cells or vascular tissues transport water, minerals, and fixed carbon throughout the organism.

Ferns and Other Seedless Vascular Plants

By the late Devonian period, plants had evolved vascular tissue, well-defined leaves, and root systems. With these advantages, plants increased in height and size. During the Carboniferous period, swamp forests of club mosses and horsetails—some specimens reaching heights of more than 30 m (100 ft)—covered most of the land. These forests gave rise to the extensive coal deposits that gave the Carboniferous its name. In seedless vascular plants, the sporophyte became the dominant phase of the lifecycle.

Water is still required for fertilization of seedless vascular plants, and most favor a moist environment. Modern-day seedless tracheophytes include club mosses, horsetails, ferns, and whisk ferns.

Phylum Lycopodiophyta: Club Mosses

The **club mosses**, or phylum **Lycopodiophyta**, are the earliest group of seedless vascular plants. They dominated the landscape of the Carboniferous, growing into tall trees and forming large swamp forests. Today's club mosses are diminutive, evergreen plants consisting of a stem (which may be branched) and microphylls ([link]). The phylum Lycopodiophyta consists of close to 1,200 species, including the quillworts (*Isoetales*), the club mosses (*Lycopodiales*), and spike mosses (*Selaginellales*), none of which are true mosses or bryophytes.

Lycophytes follow the pattern of alternation of generations seen in the bryophytes, except that the sporophyte is the major stage of the lifecycle. The gametophytes do not depend on the sporophyte for nutrients. Some gametophytes develop underground and form mycorrhizal associations with fungi. In club mosses, the sporophyte gives rise to sporophylls arranged in strobili, cone-like structures that give the class its name. Lycophytes can be homosporous or heterosporous.

In the club mosses such as *Lycopodium clavatum*, sporangia are arranged in clusters called strobili. (credit: Cory Zanker)

Phylum Monilophyta: Class Equisetopsida (Horsetails)

Horsetails, whisk ferns and ferns belong to the phylum Monilophyta, with **horsetails** placed in the Class Equisetopsida. The single genus *Equisetum* is the survivor of a large group of plants, known as Arthrophyta, which produced large trees and entire swamp forests in the Carboniferous. The plants are usually found in damp environments and marshes ([link]).

Horsetails thrive in a marsh. (credit: Myriam Feldman)

The stem of a horsetail is characterized by the presence of joints or nodes, hence the name Arthrophyta (arthro- = "joint"; -phyta = "plant"). Leaves and branches come out as whorls from the evenly spaced joints. The needle-shaped leaves do not contribute greatly to photosynthesis, the majority of which takes place in the green stem ([link]).

Thin leaves originating at the joints are noticeable on the horsetail plant. Horsetails were once used as scrubbing brushes and were nicknamed scouring brushes. (credit: Myriam Feldman)

Silica collects in the epidermal cells, contributing to the stiffness of horsetail plants. Underground stems known as rhizomes anchor the plants to the ground. Modern-day horsetails are homosporous and produce bisexual gametophytes.

Phylum Monilophyta: Class Psilotopsida (Whisk Ferns)

While most ferns form large leaves and branching roots, the **whisk ferns**, Class Psilotopsida, lack both roots and leaves, probably lost by reduction.

Photosynthesis takes place in their green stems, and small yellow knobs form at the tip of the branch stem and contain the sporangia. Whisk ferns were considered an early pterophytes. However, recent comparative DNA analysis suggests that this group may have lost both vascular tissue and roots through evolution, and is more closely related to ferns.

The whisk fern *Psilotum nudum* has conspicuous green stems with knob-shaped sporangia. (credit: Forest & Kim Starr)

Phylum Monilophyta: Class Psilotopsida (Ferns)

With their large fronds, **ferns** are the most readily recognizable seedless vascular plants. They are considered the most advanced seedless vascular plants and display characteristics commonly observed in seed plants. More than 20,000 species of ferns live in environments ranging from tropics to temperate forests. Although some species survive in dry environments, most ferns are restricted to moist, shaded places. Ferns made their appearance in the fossil record during the Devonian period and expanded during the Carboniferous.

The dominant stage of the lifecycle of a fern is the sporophyte, which consists of large compound leaves called fronds. Fronds fulfill a double role; they are photosynthetic organs that also carry reproductive organs. The stem may be buried underground as a rhizome, from which adventitious roots grow to absorb water and nutrients from the soil; or, they may grow above ground as a trunk in tree ferns ([link]). **Adventitious** organs are those that grow in unusual places, such as roots growing from the side of a stem.

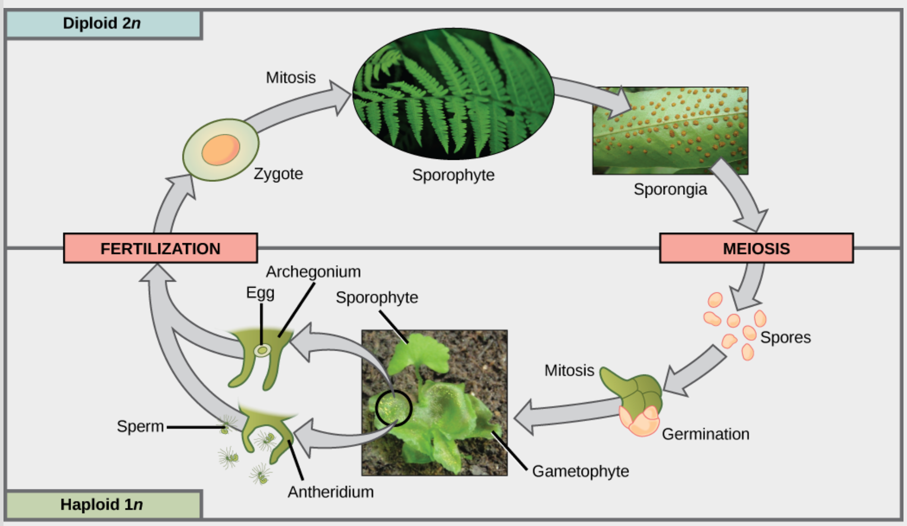
Some specimens of this short tree-fern species can grow very tall. (credit: Adrian Pingstone)

The tip of a developing fern frond is rolled into a crozier, or fiddlehead ([link]**a** and [link]**b**). Fiddleheads unroll as the frond develops.

Croziers, or fiddleheads, are the tips of fern fronds. (credit a: modification of work by Cory Zanker; credit b: modification of work by Myriam Feldman)

The lifecycle of a fern is depicted in [link].

Note:
Art Connection



This life cycle of a fern shows alternation of generations with a dominant sporophyte stage. (credit "fern": modification of work by Cory Zanker; credit "gametophyte": modification of work by "Vlmastra"/Wikimedia Commons)

Which of the following statements about the fern life cycle is false?

a. Sporangia produce haploid spores.
b. The sporophyte grows from a gametophyte.

c. The sporophyte is diploid and the gametophyte is haploid.
d. Sporangia form on the underside of the gametophyte.

Note:
Link to Learning

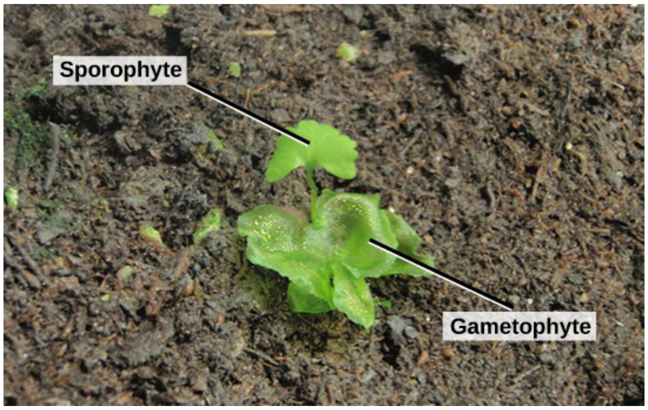
To see an animation of the lifecycle of a fern and to test your knowledge, go to the website.

Most ferns produce the same type of spores and are therefore homosporous. The diploid sporophyte is the most conspicuous stage of the lifecycle. On the underside of its mature fronds, sori (singular, sorus) form as small clusters where sporangia develop ([link]).

Sori appear as small bumps on the underside of a fern frond. (credit:

Myriam Feldman)

Inside the sori, spores are produced by meiosis and released into the air. Those that land on a suitable substrate germinate and form a heart-shaped gametophyte, which is attached to the ground by thin filamentous rhizoids ([link]).



Shown here are a young sporophyte (upper part of image) and a heart-shaped gametophyte (bottom part of image). (credit: modification of work by "Vlmastra"/Wikimedia Commons)

The inconspicuous gametophyte harbors both sex gametangia. Flagellated sperm released from the antheridium swim on a wet surface to the archegonium, where the egg is fertilized. The newly formed zygote grows into a sporophyte that emerges from the gametophyte and grows by mitosis into the next generation sporophyte.

Note:

Career Connection

Landscape Designer

Looking at the well-laid parterres of flowers and fountains in the grounds of royal castles and historic houses of Europe, it’s clear that the gardens’ creators knew about more than art and design. They were also familiar with the biology of the plants they chose. Landscape design also has strong roots in the United States’ tradition. A prime example of early American classical design is Monticello: Thomas Jefferson’s private estate. Among his many interests, Jefferson maintained a strong passion for botany. Landscape layout can encompass a small private space, like a backyard garden; public gathering places, like Central Park in New York City; or an entire city plan, like Pierre L’Enfant’s design for Washington, DC.

A landscape designer will plan traditional public spaces—such as botanical gardens, parks, college campuses, gardens, and larger developments—as well as natural areas and private gardens. The restoration of natural places encroached on by human intervention, such as wetlands, also requires the expertise of a landscape designer.

With such an array of necessary skills, a landscape designer’s education includes a solid background in botany, soil science, plant pathology, entomology, and horticulture. Coursework in architecture and design software is also required for the completion of the degree. The successful design of a landscape rests on an extensive knowledge of plant growth requirements, such as light and shade, moisture levels, compatibility of different species, and susceptibility to pathogens and pests. Mosses and ferns will thrive in a shaded area, where fountains provide moisture; cacti, on the other hand, would not fare well in that environment. The future growth of individual plants must be taken into account, to avoid crowding and competition for light and nutrients. The appearance of the space over time is also of concern. Shapes, colors, and biology must be balanced for a well-maintained and sustainable green space. Art, architecture, and biology blend in a beautifully designed and implemented landscape.

This landscaped border at a college campus was designed by students in the horticulture and landscaping department of the college. (credit: Myriam Feldman)

The Importance of Seedless Vascular Plants

Mosses and liverworts are often the first macroscopic organisms to colonize an area, both in a primary succession—where bare land is settled for the first time by living organisms—or in a secondary succession, where soil remains intact after a catastrophic event wipes out many existing species. Their spores are carried by the wind, birds, or insects. Once mosses and liverworts are established, they provide food and shelter for other species. In a hostile environment, like the tundra where the soil is frozen, bryophytes grow well because they do not have roots and can dry and rehydrate rapidly once water is again available. Mosses are at the base of the food chain in the tundra biome. Many species—from small insects to musk oxen and reindeer—depend on mosses for food. In turn, predators feed on the herbivores, which are the primary consumers. Some reports indicate that bryophytes make the soil more amenable to colonization by other plants. Because they establish symbiotic relationships with nitrogen-fixing cyanobacteria, mosses replenish the soil with nitrogen.

At the end of the nineteenth century, scientists observed that lichens and mosses were becoming increasingly rare in urban and suburban areas. Since bryophytes have neither a root system for absorption of water and nutrients, nor a cuticle layer that protects them from desiccation, pollutants in rainwater readily penetrate their tissues; they absorb moisture and nutrients through their entire exposed surfaces. Therefore, pollutants dissolved in rainwater penetrate plant tissues readily and have a larger impact on mosses than on other plants. The disappearance of mosses can be considered a bioindicator for the level of pollution in the environment.

Ferns contribute to the environment by promoting the weathering of rock, accelerating the formation of topsoil, and slowing down erosion by spreading rhizomes in the soil. The water ferns of the genus *Azolla* harbor nitrogen-fixing cyanobacteria and restore this important nutrient to aquatic habitats.

Seedless plants have historically played a role in human life through uses as tools, fuel, and medicine. Dried **peat moss**, *Sphagnum*, is commonly used as fuel in some parts of Europe and is considered a renewable resource. *Sphagnum* bogs ([link]) are cultivated with cranberry and blueberry bushes. The ability of *Sphagnum* to hold moisture makes the moss a common soil conditioner. Florists use blocks of *Sphagnum* to maintain moisture for floral arrangements.

Sphagnum acutifolium is dried peat moss and can be used as

fuel. (credit: Ken Goulding)

The attractive fronds of ferns make them a favorite ornamental plant. Because they thrive in low light, they are well suited as house plants. More importantly, fiddleheads are a traditional spring food of Native Americans in the Pacific Northwest, and are popular as a side dish in French cuisine. The licorice fern, *Polypodium glycyrrhiza*, is part of the diet of the Pacific Northwest coastal tribes, owing in part to the sweetness of its rhizomes. It has a faint licorice taste and serves as a sweetener. The rhizome also figures in the pharmacopeia of Native Americans for its medicinal properties and is used as a remedy for sore throat.

Note:
LInk to Learning

Go to this website to learn how to identify fern species based upon their fiddleheads.

By far the greatest impact of seedless vascular plants on human life, however, comes from their extinct progenitors. The tall club mosses, horsetails, and tree-like ferns that flourished in the swampy forests of the Carboniferous period gave rise to large deposits of coal throughout the world. Coal provided an abundant source of energy during the Industrial Revolution, which had tremendous consequences on human societies, including rapid technological progress and growth of large cities, as well as

the degradation of the environment. Coal is still a prime source of energy and also a major contributor to global warming.

Section Summary

Vascular systems consist of xylem tissue, which transports water and minerals, and phloem tissue, which transports sugars and proteins. With the development of the vascular system, there appeared leaves to act as large photosynthetic organs, and roots to access water from the ground. Small uncomplicated leaves are microphylls. Large leaves with vein patterns are megaphylls. Modified leaves that bear sporangia are sporophylls. Some sporophylls are arranged in cone structures called strobili.

The seedless vascular plants include club mosses, which are the most primitive; whisk ferns, which lost leaves and roots by reductive evolution; and horsetails and ferns. Ferns are the most advanced group of seedless vascular plants. They are distinguished by large leaves called fronds and small sporangia-containing structures called sori, which are found on the underside of the fronds.

Mosses play an essential role in the balance of the ecosystems; they are pioneering species that colonize bare or devastated environments and make it possible for a succession to occur. They contribute to the enrichment of the soil and provide shelter and nutrients for animals in hostile environments. Mosses and ferns can be used as fuels and serve culinary, medical, and decorative purposes.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the fern life cycle is false?

a. Sporangia produce haploid spores.

b. The sporophyte grows from a gametophyte.
c. The sporophyte is diploid and the gametophyte is haploid.
d. Sporangia form on the underside of the gametophyte.

Solution:

[link] D.

Review Questions

Exercise:

Problem: Microphylls are characteristic of which types of plants?

a. mosses
b. liverworts
c. club mosses
d. ferns

Solution:

C

Exercise:

Problem:

A plant in the understory of a forest displays a segmented stem and slender leaves arranged in a whorl. It is probably a _________.

a. club moss
b. whisk fern
c. fern
d. horsetail

Solution:

D

Exercise:

Problem:

The following structures are found on the underside of fern leaves and contain sporangia:

a. sori
b. rhizomes
c. megaphylls
d. microphylls

Solution:

A

Exercise:

Problem: The dominant organism in fern is the _________.

a. sperm
b. spore
c. gamete
d. sporophyte

Solution:

D

Exercise:

Problem: What seedless plant is a renewable source of energy?

a. club moss
b. horsetail
c. sphagnum moss
d. fern

Solution:

C

Exercise:

Problem: How do mosses contribute to returning nitrogen to the soil?

a. Mosses fix nitrogen from the air.
b. Mosses harbor cyanobacteria that fix nitrogen.
c. Mosses die and return nitrogen to the soil.
d. Mosses decompose rocks and release nitrogen.

Solution:

D

Free Response

Exercise:

Problem:

How did the development of a vascular system contribute to the increase in size of plants?

Solution:

Plants became able to transport water and nutrients and not be limited by rates of diffusion. Vascularization allowed the development of

leaves, which increased efficiency of photosynthesis and provided more energy for plant growth.

Exercise:

Problem:

Which plant is considered the most advanced seedless vascular plant and why?

Solution:

Ferns are considered the most advanced seedless vascular plants, because they display characteristics commonly observed in seed plants—they form large leaves and branching roots.

Glossary

adventitious
: describes an organ that grows in an unusual place, such as a roots growing from the side of a stem

club mosses
: earliest group of seedless vascular plants

fern
: seedless vascular plant that produces large fronds; the most advanced group of seedless vascular plants

horsetail
: seedless vascular plant characterized by joints

lignin
: complex polymer impermeable to water

lycophyte
: club moss

megaphyll
larger leaves with a pattern of branching veins

microphyll
small size and simple vascular system with a single unbranched vein

peat moss
Sphagnum

phloem
tissue responsible for transport of sugars, proteins, and other solutes

sporophyll
leaf modified structurally to bear sporangia

strobili
cone-like structures that contain the sporangia

tracheophyte
vascular plant

vein
bundle of vascular tissue made of xylem and phloem

whisk fern
seedless vascular plant that lost roots and leaves by reduction

xylem
tissue responsible for long-distance transport of water and nutrients

Gymnosperms

By the end of this section, you will be able to:

- Discuss the type of seeds produced by gymnosperms, as well as other characteristics of gymnosperms
- State which period saw the first appearance of gymnosperms and explain when they were the dominant plant life
- List the four groups of modern-day gymnosperms and provide examples of each

Gymnosperms, meaning “naked seeds,” are a diverse group of seed plants and are paraphyletic. Paraphyletic groups are those in which not all members are descendants of a single common ancestor. Their characteristics include naked seeds, separate female and male gametes, pollination by wind, and tracheids (which transport water and solutes in the vascular system).

Gymnosperm seeds are not enclosed in an ovary; rather, they are exposed on cones or modified leaves. Sporophylls are specialized leaves that produce sporangia. The term **strobilus** (plural = strobili) describes a tight arrangement of sporophylls around a central stalk, as seen in cones. Some seeds are enveloped by sporophyte tissues upon maturation. The layer of sporophyte tissue that surrounds the megasporangium, and later, the embryo, is called the **integument**.

Gymnosperms were the dominant phylum in Mesozoic era. They are adapted to live where fresh water is scarce during part of the year, or in the nitrogen-poor soil of a bog. Therefore, they are still the prominent phylum in the coniferous biome or taiga, where the evergreen conifers have a selective advantage in cold and dry weather. Evergreen conifers continue low levels of photosynthesis during the cold months, and are ready to take advantage of the first sunny days of spring. One disadvantage is that conifers are more susceptible than deciduous trees to infestations because conifers do not lose their leaves all at once. They cannot, therefore, shed parasites and restart with a fresh supply of leaves in spring.

The life cycle of a gymnosperm involves alternation of generations, with a dominant sporophyte in which the female gametophyte resides, and reduced

gametophytes. All gymnosperms are heterosporous. The male and female reproductive organs can form in cones or strobili. Male and female sporangia are produced either on the same plant, described as **monoecious** (“one home” or bisexual), or on separate plants, referred to as **dioecious** (“two homes” or unisexual) plants. The life cycle of a conifer will serve as our example of reproduction in gymnosperms.

Diversity of Gymnosperms

Modern gymnosperms are classified into four phyla. Coniferophyta, Cycadophyta, and Ginkgophyta are similar in their production of secondary cambium (cells that generate the vascular system of the trunk or stem and are partially specialized for water transportation) and their pattern of seed development. However, the three phyla are not closely related phylogenetically to each other. Gnetophyta are considered the closest group to angiosperms because they produce true xylem tissue.

Conifers (Coniferophyta)

Conifers are the dominant phylum of gymnosperms, with the most variety of species ([link]). Most are typically tall trees that usually bear scale-like or needle-like leaves. Water evaporation from leaves is reduced by their thin shape and the thick cuticle. Snow slides easily off needle-shaped leaves, keeping the load light and decreasing breaking of branches. Adaptations to cold and dry weather explain the predominance of conifers at high altitudes and in cold climates. Conifers include familiar evergreen trees such as pines, spruces, firs, cedars, sequoias, and yews. A few species are deciduous and lose their leaves in fall. The European larch and the tamarack are examples of deciduous conifers ([link]**c**). Many coniferous trees are harvested for paper pulp and timber. The wood of conifers is more primitive than the wood of angiosperms; it contains tracheids, but no vessel elements, and is therefore referred to as “soft wood.”

Conifers are the dominant form of vegetation in cold or arid environments and at high altitudes. Shown here are the (a) evergreen spruce *Picea* sp., (b) juniper *Juniperus* sp., (c) sequoia *Sequoia Semervirens*, which is a deciduous gymnosperm, and (d) the tamarack *Larix larcinia*. Notice the yellow leaves of the tamarack. (credit a: modification of work by Rosendahl; credit b: modification of work by Alan Levine; credit c: modification of work by Wendy McCormic; credit d: modification of work by Micky Zlimen)

Cycads

Cycads thrive in mild climates, and are often mistaken for palms because of the shape of their large, compound leaves. Cycads bear large cones ([link]), and may be pollinated by beetles rather than wind: unusual for a gymnosperm. They dominated the landscape during the age of dinosaurs in the Mesozoic, but only a hundred or so species persisted to modern times. They face possible extinction, and several species are protected through international conventions. Because of their attractive shape, they are often used as ornamental plants in gardens in the tropics and subtropics.

This *Encephalartos ferox* cycad has large cones and broad, fern-like leaves. (credit: Wendy Cutler)

Gingkophytes

The single surviving species of the **gingkophytes** group is the *Gingko biloba* ([link]). Its fan-shaped leaves—unique among seed plants because they feature a dichotomous venation pattern—turn yellow in autumn and fall from the tree. For centuries, *G. biloba* was cultivated by Chinese Buddhist monks in monasteries, which ensured its preservation. It is planted

in public spaces because it is unusually resistant to pollution. Male and female organs are produced on separate plants. Typically, gardeners plant only male trees because the seeds produced by the female plant have an off-putting smell of rancid butter.

This plate from the 1870 book *Flora Japonica*, *Sectio Prima (Tafelband)* depicts the leaves and fruit of *Gingko biloba*, as drawn by Philipp Franz von Siebold and Joseph Gerhard Zuccarini.

Gnetophytes

Gnetophytes are the closest relative to modern angiosperms, and include three dissimilar genera of plants: *Ephedra*, *Gnetum*, and *Welwitschia* ([link]). Like angiosperms, they have broad leaves. In tropical and subtropical zones, gnetophytes are vines or small shrubs. *Ephedra* occurs in dry areas of the West Coast of the United States and Mexico. *Ephedra*'s small, scale-like leaves are the source of the compound ephedrine, which is used in medicine as a potent decongestant. Because ephedrine is similar to amphetamines, both in chemical structure and neurological effects, its use is restricted to prescription drugs. Like angiosperms, but unlike other gymnosperms, all gnetophytes possess vessel elements in their xylem.

(a) *Ephedra viridis*, known by the common name *Mormon tea*, grows on the West Coast of the United States and Mexico. (b) *Gnetum gnemon* grows in Malaysia. (c) The large *Welwitschia mirabilis* can be found in the Namibian desert. (credit a: modification of work by USDA; credit b: modification of work by Malcolm Manners; credit c: modification of work by Derek Keats)

Note:
Link to Learning

Watch this BBC video describing the amazing strangeness of Welwitschia.

Section Summary

Gymnosperms are heterosporous seed plants that produce naked seeds. They appeared in the Paleozoic period and were the dominant plant life during the Mesozoic. Modern-day gymnosperms belong to four phyla. The largest phylum, Coniferophyta, is represented by conifers, the predominant plants at high altitude and latitude. Cycads (phylum Cycadophyta) resemble palm trees and grow in tropical climates. *Gingko biloba* is the only representative of the phylum Gingkophyta. The last phylum, Gnetophyta, is a diverse group of shrubs that produce vessel elements in their wood.

Art Connections

Exercise:

Problem:[link] At what stage does the diploid zygote form?

a. When the female cone begins to bud from the tree
b. At fertilization
c. When the seeds drop from the tree
d. When the pollen tube begins to grow

Solution:

[link] B. The diploid zygote forms after the pollen tube has finished forming, so that the male generative nuclei can fuse with the female

gametophyte.

Review Questions

Exercise:

Problem: Which of the following traits characterizes gymnosperms?

a. The plants carry exposed seeds on modified leaves.
b. Reproductive structures are located in a flower.
c. After fertilization, the ovary thickens and forms a fruit.
d. The gametophyte is longest phase of the life cycle.

Solution:

A

Exercise:

Problem:

Megasporocytes will eventually produce which of the following?

a. pollen grain
b. sporophytes
c. male gametophytes
d. female gametophytes

Solution:

D

Exercise:

Problem:

What is the ploidy of the following structures: gametophyte, seed, spore, sporophyte?

a. 1*n*, 1*n*, 2*n*, 2*n*
b. 1*n*, 2*n*, 1*n*, 2*n*
c. 2*n*, 1*n*, 2*n*, 1*n*
d. 2*n*, 2*n*, 1*n*, 1*n*

Solution:

B

Exercise:

Problem: In the northern forests of Siberia, a tall tree is most likely a:

a. conifer
b. cycad
c. *Gingko biloba*
d. gnetophyte

Solution:

A

Free Response

Exercise:

Problem:

The Mediterranean landscape along the sea shore is dotted with pines and cypresses. The weather is not cold, and the trees grow at sea level. What evolutionary adaptation of conifers makes them suitable to the Mediterranean climate?

Solution:

The trees are adapted to arid weather, and do not lose as much water due to transpiration as non-conifers.

Exercise:

Problem: What are the four modern-day phyla of gymnosperms?

Solution:

The four modern-day phyla of gymnosperms are Coniferophyta, Cycadophyta, Gingkophyta, and Gnetophyta.

Glossary

conifer
: dominant phylum of gymnosperms with the most variety of trees

cycad
: gymnosperm that grows in tropical climates and resembles a palm tree; member of the phylum Cycadophyta

dioecious
: describes a species in which the male and female reproductive organs are carried on separate specimens

gingkophyte
: gymnosperm with one extant species, the *Gingko biloba*: a tree with fan-shaped leaves

gnetophyte
: gymnosperm shrub with varied morphological features that produces vessel elements in its woody tissues; the phylum includes the genera *Ephedra, Gnetum* and *Welwitschia*

gymnosperm
: seed plant with naked seeds (seeds exposed on modified leaves or in cones)

integument
: layer of sporophyte tissue that surrounds the megasporangium, and later, the embryo

megasporocyte
: megaspore mother cell; larger spore that germinates into a female gametophyte in a heterosporous plant

microsporocyte
: smaller spore that produces a male gametophyte in a heterosporous plant

monoecious
: describes a species in which the male and female reproductive organs are on the same plant

ovulate cone
: cone containing two ovules per scale

strobilus
: plant structure with a tight arrangement of sporophylls around a central stalk, as seen in cones or flowers; the male strobilus produces pollen, and the female strobilus produces eggs

Angiosperms
By the end of this section, you will be able to:

- Explain why angiosperms are the dominant form of plant life in most terrestrial ecosystems
- Discuss the two main groups of flowering plants

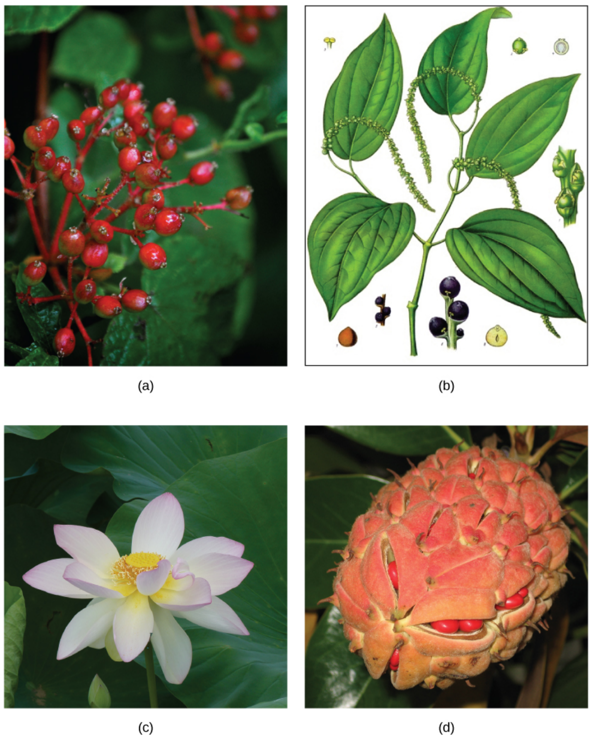
Diversity of Angiosperms

Angiosperms are classified in a single phylum: the **Anthophyta**. Modern angiosperms appear to be a monophyletic group, which means that they originate from a single ancestor. Flowering plants are divided into two major groups, according to the structure of the cotyledons, pollen grains, and other structures. **Monocots** include grasses and lilies, and eudicots or **dicots** form a polyphyletic group. **Basal angiosperms** are a group of plants that are believed to have branched off before the separation into monocots and eudicots because they exhibit traits from both groups. They are categorized separately in many classification schemes. The *Magnoliidae* (magnolia trees, laurels, and water lilies) and the *Piperaceae* (peppers) belong to the basal angiosperm group.

Basal Angiosperms

The Magnoliidae are represented by the magnolias: tall trees bearing large, fragrant flowers that have many parts and are considered archaic ([link]**d**). Laurel trees produce fragrant leaves and small, inconspicuous flowers. The *Laurales* grow mostly in warmer climates and are small trees and shrubs. Familiar plants in this group include the bay laurel, cinnamon, spice bush ([link]**a**), and avocado tree. The *Nymphaeales* are comprised of the water lilies, lotus ([link]**c**), and similar plants; all species thrive in freshwater biomes, and have leaves that float on the water surface or grow underwater. Water lilies are particularly prized by gardeners, and have graced ponds and pools for thousands of years. The *Piperales* are a group of herbs, shrubs, and small trees that grow in the tropical climates. They have small flowers without petals that are tightly arranged in long spikes. Many species are the

source of prized fragrance or spices, for example the berries of *Piper nigrum* ([link]**b**) are the familiar black peppercorns that are used to flavor many dishes.



The (a) common spicebush belongs to the *Laurales*, the same family as cinnamon and bay laurel. The fruit of (b) the *Piper nigrum* plant is black pepper, the main product that was traded along spice routes. Notice the small, unobtrusive, clustered flowers. (c) Lotus flowers, *Nelumbo nucifera*, have been cultivated since ancient times for their ornamental value; the root of the lotus flower is eaten as a vegetable. The red seeds of (d) a magnolia tree, characteristic of the

final stage, are just starting to appear. (credit a: modification of work by Cory Zanker; credit b: modification of work by Franz Eugen Köhler; credit c: modification of work by "berduchwal"/Flickr; credit d: modification of work by "Coastside2"/Wikimedia Commons).

Monocots

Plants in the monocot group are primarily identified as such by the presence of a single cotyledon in the seedling. Other anatomical features shared by monocots include veins that run parallel to the length of the leaves, and flower parts that are arranged in a three- or six-fold symmetry. True woody tissue is rarely found in monocots. In palm trees, vascular and parenchyma tissues produced by the primary and secondary thickening meristems form the trunk. The pollen from the first angiosperms was monosulcate, containing a single furrow or pore through the outer layer. This feature is still seen in the modern monocots. Vascular tissue of the stem is not arranged in any particular pattern. The root system is mostly adventitious and unusually positioned, with no major tap root. The monocots include familiar plants such as the true lilies (which are at the origin of their alternate name of Liliopsida), orchids, grasses, and palms. Many important crops are monocots, such as rice and other cereals, corn, sugar cane, and tropical fruits like bananas and pineapples ([link]).

The world’s major crops are flowering plants. (a) Rice, (b) wheat, and (c) bananas are monocots, while (d) cabbage, (e) beans, and (f) peaches are dicots. (credit a: modification of work by David Nance, USDA ARS; credit b, c: modification of work by Rosendahl; credit d: modification of work by Bill Tarpenning, USDA; credit e: modification of work by Scott Bauer, USDA ARS; credit f: modification of work by Keith Weller, USDA)

Eudicots

Eudicots, or true dicots, are characterized by the presence of two cotyledons in the developing shoot. Veins form a network in leaves, and flower parts come in four, five, or many whorls. Vascular tissue forms a ring in the stem; in monocots, vascular tissue is scattered in the stem. Eudicots can be **herbaceous** (like grasses), or produce woody tissues. Most eudicots produce pollen that is trisulcate or triporate, with three furrows or pores.

The root system is usually anchored by one main root developed from the embryonic radicle. Eudicots comprise two-thirds of all flowering plants. The major differences between monocots and eudicots are summarized in [link]. Many species exhibit characteristics that belong to either group; as such, the classification of a plant as a monocot or a eudicot is not always clearly evident.

Comparison of Structural Characteristics of Monocots and Eudicots		
Characteristic	**Monocot**	**Eudicot**
Cotyledon	One	Two
Veins in Leaves	Parallel	Network (branched)
Stem Vascular Tissue	Scattered	Arranged in ring pattern
Roots	Network of adventitious roots	Tap root with many lateral roots
Pollen	Monosulcate	Trisulcate
Flower Parts	Three or multiple of three	Four, five, multiple of four or five and whorls

Section Summary

Angiosperms are the dominant form of plant life in most terrestrial ecosystems, comprising about 90 percent of all plant species. Most crops and ornamental plants are angiosperms. Their success comes from two innovative structures that protect reproduction from variability in the environment: the flower and the fruit. Flowers were derived from modified leaves. The main parts of a flower are the sepals and petals, which protect the reproductive parts: the stamens and the carpels. The stamens produce the male gametes in pollen grains. The carpels contain the female gametes (the eggs inside the ovules), which are within the ovary of a carpel. The walls of the ovary thicken after fertilization, ripening into fruit that ensures dispersal by wind, water, or animals.

The angiosperm life cycle is dominated by the sporophyte stage. Double fertilization is an event unique to angiosperms. One sperm in the pollen fertilizes the egg, forming a diploid zygote, while the other combines with the two polar nuclei, forming a triploid cell that develops into a food storage tissue called the endosperm. Flowering plants are divided into two main groups, the monocots and eudicots, according to the number of cotyledons in the seedlings. Basal angiosperms belong to an older lineage than monocots and eudicots.

Art Connections

Exercise:

Problem:

[link] If a flower lacked a megasporangium, what type of gamete would not form? If the flower lacked a microsporangium, what type of gamete would not form?

Solution:

[link] Without a megasporangium, an egg would not form; without a microsporangium, pollen would not form.

Review Questions

Exercise:

Problem:

Which of the following structures in a flower is not directly involved in reproduction?

a. the style
b. the stamen
c. the sepal
d. the anther

Solution:

C

Exercise:

Problem: Pollen grains develop in which structure?

a. the anther
b. the stigma
c. the filament
d. the carpel

Solution:

A

Exercise:

Problem:

In the course of double fertilization, one sperm cell fuses with the egg and the second one fuses with _________.

a. the synergids
b. the polar nuclei of the center cell

c. the egg as well
d. the antipodal cells

Solution:

B

Exercise:

Problem:

Corn develops from a seedling with a single cotyledon, displays parallel veins on its leaves, and produces monosulcate pollen. It is most likely:

a. a gymnosperm
b. a monocot
c. a eudicot
d. a basal angiosperm

Solution:

B

Free Response

Exercise:

Problem:

Some cycads are considered endangered species and their trade is severely restricted. Customs officials stop suspected smugglers who claim that the plants in their possession are palm trees, not cycads. How would a botanist distinguish between the two types of plants?

Solution:

The resemblance between cycads and palm trees is only superficial. Cycads are gymnosperms and do not bear flowers or fruit. Cycads produce cones: large, female cones that produce naked seeds, and smaller male cones on separate plants. Palms do not.

Exercise:

Problem:

What are the two structures that allow angiosperms to be the dominant form of plant life in most terrestrial ecosystems?

Solution:

Angiosperms are successful because of flowers and fruit. These structures protect reproduction from variability in the environment.

Glossary

anther
: sac-like structure at the tip of the stamen in which pollen grains are produced

Anthophyta
: phylum to which angiosperms belong

basal angiosperms
: a group of plants that probably branched off before the separation of monocots and eudicots

calyx
: whorl of sepals

carpel
: single unit of the pistil

corolla
: collection of petals

cotyledon
    primitive leaf that develop in the zygote; monocots have one cotyledon, and dicots have two cotyledons

dicot
    (also, eudicot) related group of angiosperms whose embryos possess two cotyledons

filament
    thin stalk that links the anther to the base of the flower

gynoecium
    (also, carpel) structure that constitute the female reproductive organ

herbaceous
    grass-like plant noticeable by the absence of woody tissue

monocot
    related group of angiosperms that produce embryos with one cotyledon and pollen with a single ridge

ovary
    chamber that contains and protects the ovule or female megasporangium

perianth
    part of the plant consisting of the calyx (sepals) and corolla (petals)

petal
    modified leaf interior to the sepals; colorful petals attract animal pollinators

pistil
    fused group of carpels

sepal
    modified leaf that encloses the bud; outermost structure of a flower

stamen

structure that contains the male reproductive organs

stigma
: uppermost structure of the carpel where pollen is deposited

style
: long, thin structure that links the stigma to the ovary

The Role of Seed Plants
By the end of this section, you will be able to:

- Explain how angiosperm diversity is due, in part, to multiple interactions with animals
- Describe ways in which pollination occurs
- Discuss the roles that plants play in ecosystems and how deforestation threatens plant biodiversity

Without seed plants, life as we know it would not be possible. Plants play a key role in the maintenance of terrestrial ecosystems through stabilization of soils, cycling of carbon, and climate moderation. Large tropical forests release oxygen and act as carbon dioxide sinks. Seed plants provide shelter to many life forms, as well as food for herbivores, thereby indirectly feeding carnivores. Plant secondary metabolites are used for medicinal purposes and industrial production.

Animals and Plants: Herbivory

Coevolution of flowering plants and insects is a hypothesis that has received much attention and support, especially because both angiosperms and insects diversified at about the same time in the middle Mesozoic. Many authors have attributed the diversity of plants and insects to pollination and **herbivory**, or consumption of plants by insects and other animals. This is believed to have been as much a driving force as pollination. Coevolution of herbivores and plant defenses is observed in nature. Unlike animals, most plants cannot outrun predators or use mimicry to hide from hungry animals. A sort of arms race exists between plants and herbivores. To “combat” herbivores, some plant seeds—such as acorn and unripened persimmon—are high in alkaloids and therefore unsavory to some animals. Other plants are protected by bark, although some animals developed specialized mouth pieces to tear and chew vegetal material. Spines and thorns ([link]) deter most animals, except for mammals with thick fur, and some birds have specialized beaks to get past such defenses.

(a) Spines and (b) thorns are examples of plant defenses. (credit a: modification of work by Jon Sullivan; credit b: modification of work by I. Sáček, Sr.)

Herbivory has been used by seed plants for their own benefit in a display of mutualistic relationships. The dispersal of fruit by animals is the most striking example. The plant offers to the herbivore a nutritious source of food in return for spreading the plant’s genetic material to a wider area.

An extreme example of collaboration between an animal and a plant is the case of acacia trees and ants. The trees support the insects with shelter and food. In return, ants discourage herbivores, both invertebrates and vertebrates, by stinging and attacking leaf-eating insects.

Animals and Plants: Pollination

Grasses are a successful group of flowering plants that are wind pollinated. They produce large amounts of powdery pollen carried over large distances by the wind. The flowers are small and wisp-like. Large trees such as oaks, maples, and birches are also wind pollinated.

Note:
Link to Learning

Explore this website for additional information on pollinators.

More than 80 percent of angiosperms depend on animals for **pollination**: the transfer of pollen from the anther to the stigma. Consequently, plants have developed many adaptations to attract pollinators. The specificity of specialized plant structures that target animals can be very surprising. It is possible, for example, to determine the type of pollinator favored by a plant just from the flower's characteristics. Many bird or insect-pollinated flowers secrete **nectar**, which is a sugary liquid. They also produce both fertile pollen, for reproduction, and sterile pollen rich in nutrients for birds and insects. Butterflies and bees can detect ultraviolet light. Flowers that attract these pollinators usually display a pattern of low ultraviolet reflectance that helps them quickly locate the flower's center and collect nectar while being dusted with pollen ([link]). Large, red flowers with little smell and a long funnel shape are preferred by hummingbirds, who have good color perception, a poor sense of smell, and need a strong perch. White flowers opened at night attract moths. Other animals—such as bats, lemurs, and lizards—can also act as pollinating agents. Any disruption to these interactions, such as the disappearance of bees as a consequence of colony collapse disorders, can lead to disaster for agricultural industries that depend heavily on pollinated crops.

As a bee collects nectar from a flower, it is dusted by pollen, which it then disperses to other flowers. (credit: John Severns)

Note:
Scientific Method Connection
Testing Attraction of Flies by Rotting Flesh Smell
Question: Will flowers that offer cues to bees attract carrion flies if sprayed with compounds that smell like rotten flesh?
Background: Visitation of flowers by pollinating flies is a function mostly of smell. Flies are attracted by rotting flesh and carrions. The putrid odor seems to be the major attractant. The polyamines putrescine and cadaverine, which are the products of protein breakdown after animal death, are the source of the pungent smell of decaying meat. Some plants strategically attract flies by synthesizing polyamines similar to those generated by decaying flesh and thereby attract carrion flies.
Flies seek out dead animals because they normally lay their eggs on them and their maggots feed on the decaying flesh. Interestingly, time of death can be determined by a forensic entomologist based on the stages and type of maggots recovered from cadavers.

Hypothesis: Because flies are drawn to other organisms based on smell and not sight, a flower that is normally attractive to bees because of its colors will attract flies if it is sprayed with polyamines similar to those generated by decaying flesh.
Test the hypothesis:

1. Select flowers usually pollinated by bees. White petunia may be good choice.
2. Divide the flowers into two groups, and while wearing eye protection and gloves, spray one group with a solution of either putrescine or cadaverine. (Putrescine dihydrochloride is typically available in 98 percent concentration; this can be diluted to approximately 50 percent for this experiment.)
3. Place the flowers in a location where flies are present, keeping the sprayed and unsprayed flowers separated.
4. Observe the movement of the flies for one hour. Record the number of visits to the flowers using a table similar to [link]. Given the rapid movement of flies, it may be beneficial to use a video camera to record the fly–flower interaction. Replay the video in slow motion to obtain an accurate record of the number of fly visits to the flowers.
5. Repeat the experiment four more times with the same species of flower, but using different specimens.
6. Repeat the entire experiment with a different type of flower that is normally pollinated by bees.

Results of Number of Visits by Flies to Sprayed and Control/Unsprayed Flowers

Trial #	**Sprayed Flowers**	**Unsprayed Flowers**
1		

Results of Number of Visits by Flies to Sprayed and Control/Unsprayed Flowers		
Trial #	**Sprayed Flowers**	**Unsprayed Flowers**
2		
3		
4		
5		

Analyze your data: Review the data you have recorded. Average the number of visits that flies made to sprayed flowers over the course of the five trials (on the first flower type) and compare and contrast them to the average number of visits that flies made to the unsprayed/control flowers. Can you draw any conclusions regarding the attraction of the flies to the sprayed flowers?
For the second flower type used, average the number of visits that flies made to sprayed flowers over the course of the five trials and compare and contrast them to the average number of visits that flies made to the unsprayed/control flowers. Can you draw any conclusions regarding the attraction of the flies to the sprayed flowers?
Compare and contrast the average number of visits that flies made to the two flower types. Can you draw any conclusions about whether the appearance of the flower had any impact on the attraction of flies? Did smell override any appearance differences, or were the flies attracted to one flower type more than another?
Form a conclusion: Do the results support the hypothesis? If not, how can this be explained?

The Importance of Seed Plants in Human Life

Seed plants are the foundation of human diets across the world ([link]). Many societies eat almost exclusively vegetarian fare and depend solely on seed plants for their nutritional needs. A few **crops** (rice, wheat, and potatoes) dominate the agricultural landscape. Many crops were developed during the agricultural revolution, when human societies made the transition from nomadic hunter–gatherers to horticulture and agriculture. Cereals, rich in carbohydrates, provide the staple of many human diets. Beans and nuts supply proteins. Fats are derived from crushed seeds, as is the case for peanut and rapeseed (canola) oils, or fruits such as olives. Animal husbandry also consumes large amounts of crops.

Staple crops are not the only food derived from seed plants. Fruits and vegetables provide nutrients, vitamins, and fiber. Sugar, to sweeten dishes, is produced from the monocot sugarcane and the eudicot sugar beet. Drinks are made from infusions of tea leaves, chamomile flowers, crushed coffee beans, or powdered cocoa beans. Spices come from many different plant parts: saffron and cloves are stamens and buds, black pepper and vanilla are seeds, the bark of a bush in the *Laurales* family supplies cinnamon, and the herbs that flavor many dishes come from dried leaves and fruit, such as the pungent red chili pepper. The volatile oils of flowers and bark provide the scent of perfumes. Additionally, no discussion of seed plant contribution to human diet would be complete without the mention of alcohol. Fermentation of plant-derived sugars and starches is used to produce alcoholic beverages in all societies. In some cases, the beverages are derived from the fermentation of sugars from fruit, as with wines and, in other cases, from the fermentation of carbohydrates derived from seeds, as with beers.

Seed plants have many other uses, including providing wood as a source of timber for construction, fuel, and material to build furniture. Most paper is derived from the pulp of coniferous trees. Fibers of seed plants such as cotton, flax, and hemp are woven into cloth. Textile dyes, such as indigo, were mostly of plant origin until the advent of synthetic chemical dyes.

Lastly, it is more difficult to quantify the benefits of ornamental seed plants. These grace private and public spaces, adding beauty and serenity to human lives and inspiring painters and poets alike.

(a)

(b)

(c)

(d)

Humans rely on plants for a variety of reasons. (a) Cacao beans were introduced in Europe from the New World, where they were used by Mesoamerican civilizations. Combined with sugar, another plant product, chocolate is a popular food. (b) Flowers like the tulip are cultivated for their beauty. (c) Quinine, extracted from cinchona trees, is used to treat malaria, to reduce fever, and to alleviate pain. (d) This violin is made of wood. (credit a: modification of work by "Everjean"/Flickr; credit b:

modification of work by Rosendahl; credit c: modification of work by Franz Eugen Köhler)

The medicinal properties of plants have been known to human societies since ancient times. There are references to the use of plants' curative properties in Egyptian, Babylonian, and Chinese writings from 5,000 years ago. Many modern synthetic therapeutic drugs are derived or synthesized de novo from plant secondary metabolites. It is important to note that the same plant extract can be a therapeutic remedy at low concentrations, become an addictive drug at higher doses, and can potentially kill at high concentrations. [link] presents a few drugs, their plants of origin, and their medicinal applications.

Plant Origin of Medicinal Compounds and Medical Applications		
Plant	**Compound**	**Application**
Deadly nightshade (*Atropa belladonna*)	Atropine	Dilate eye pupils for eye exams
Foxglove (*Digitalis purpurea*)	Digitalis	Heart disease, stimulates heart beat
Yam (*Dioscorea* spp.)	Steroids	Steroid hormones: contraceptive pill and cortisone

Plant Origin of Medicinal Compounds and Medical Applications		
Plant	**Compound**	**Application**
Ephedra (*Ephedra* spp.)	Ephedrine	Decongestant and bronchiole dilator
Pacific yew (*Taxus brevifolia*)	Taxol	Cancer chemotherapy; inhibits mitosis
Opium poppy (*Papaver somniferum*)	Opioids	Analgesic (reduces pain without loss of consciousness) and narcotic (reduces pain with drowsiness and loss of consciousness) in higher doses
Quinine tree (*Cinchona* spp.)	Quinine	Antipyretic (lowers body temperature) and antimalarial
Willow (*Salix* spp.)	Salicylic acid (aspirin)	Analgesic and antipyretic

Note:
Career Connection
Ethnobotanist
The relatively new field of ethnobotany studies the interaction between a particular culture and the plants native to the region. Seed plants have a large influence on day-to-day human life. Not only are plants the major source of food and medicine, they also influence many other aspects of society, from clothing to industry. The medicinal properties of plants were

recognized early on in human cultures. From the mid-1900s, synthetic chemicals began to supplant plant-based remedies.

Pharmacognosy is the branch of pharmacology that focuses on medicines derived from natural sources. With massive globalization and industrialization, there is a concern that much human knowledge of plants and their medicinal purposes will disappear with the cultures that fostered them. This is where ethnobotanists come in. To learn about and understand the use of plants in a particular culture, an ethnobotanist must bring in knowledge of plant life and an understanding and appreciation of diverse cultures and traditions. The Amazon forest is home to an incredible diversity of vegetation and is considered an untapped resource of medicinal plants; yet, both the ecosystem and its indigenous cultures are threatened with extinction.

To become an ethnobotanist, a person must acquire a broad knowledge of plant biology, ecology and sociology. Not only are the plant specimens studied and collected, but also the stories, recipes, and traditions that are linked to them. For ethnobotanists, plants are not viewed solely as biological organisms to be studied in a laboratory, but as an integral part of human culture. The convergence of molecular biology, anthropology, and ecology make the field of ethnobotany a truly multidisciplinary science.

Biodiversity of Plants

Biodiversity ensures a resource for new food crops and medicines. Plant life balances ecosystems, protects watersheds, mitigates erosion, moderates climate and provides shelter for many animal species. Threats to plant diversity, however, come from many angles. The explosion of the human population, especially in tropical countries where birth rates are highest and economic development is in full swing, is leading to human encroachment into forested areas. To feed the larger population, humans need to obtain arable land, so there is massive clearing of trees. The need for more energy to power larger cities and economic growth therein leads to the construction of dams, the consequent flooding of ecosystems, and increased emissions of pollutants. Other threats to tropical forests come from poachers, who log trees for their precious wood. Ebony and Brazilian rosewood, both on the

endangered list, are examples of tree species driven almost to extinction by indiscriminate logging.

The number of plant species becoming extinct is increasing at an alarming rate. Because ecosystems are in a delicate balance, and seed plants maintain close symbiotic relationships with animals—whether predators or pollinators—the disappearance of a single plant can lead to the extinction of connected animal species. A real and pressing issue is that many plant species have not yet been catalogued, and so their place in the ecosystem is unknown. These unknown species are threatened by logging, habitat destruction, and loss of pollinators. They may become extinct before we have the chance to begin to understand the possible impacts from their disappearance. Efforts to preserve biodiversity take several lines of action, from preserving heirloom seeds to barcoding species. **Heirloom seeds** come from plants that were traditionally grown in human populations, as opposed to the seeds used for large-scale agricultural production. **Barcoding** is a technique in which one or more short gene sequences, taken from a well-characterized portion of the genome, are used to identify a species through DNA analysis.

Section Summary

Angiosperm diversity is due in part to multiple interactions with animals. Herbivory has favored the development of defense mechanisms in plants, and avoidance of those defense mechanism in animals. Pollination (the transfer of pollen to a carpel) is mainly carried out by wind and animals, and angiosperms have evolved numerous adaptations to capture the wind or attract specific classes of animals.

Plants play a key role in ecosystems. They are a source of food and medicinal compounds, and provide raw materials for many industries. Rapid deforestation and industrialization, however, threaten plant biodiversity. In turn, this threatens the ecosystem.

Review Questions

Exercise:

Problem:

Which of the following plant structures is not a defense against herbivory?

a. thorns
b. spines
c. nectar
d. alkaloids

Solution:

C

Exercise:

Problem:

White and sweet-smelling flowers with abundant nectar are probably pollinated by

a. bees and butterflies
b. flies
c. birds
d. wind

Solution:

A

Exercise:

Problem:

Abundant and powdery pollen produced by small, indistinct flowers is probably transported by:

a. bees and butterflies

b. flies
c. birds
d. wind

Solution:

D

Exercise:

Problem: Plants are a source of ________.

a. food
b. fuel
c. medicine
d. all of the above

Solution:

D

Free Response

Exercise:

Problem:

Biosynthesis of nectar and nutrient-rich pollen is energetically very expensive for a plant. Yet, plants funnel large amounts of energy into animal pollination. What are the evolutionary advantages that offset the cost of attracting animal pollinators?

Solution:

Using animal pollinators promotes cross-pollination and increases genetic diversity. The odds that the pollen will reach another flower are

greatly increased compared with the randomness of wind pollination.

Exercise:

Problem:

What is biodiversity and why is it important to an ecosystem?

Solution:

Biodiversity is the variation in all forms of life. It can refer to variation within a species, within an ecosystem, or on an entire planet. It is important because it ensures a resource for new food crops and medicines. Plant life balances the ecosystems, protects watersheds, mitigates erosion, moderates climate, and provides shelter for many animal species.

Glossary

barcoding
: molecular biology technique in which one or more short gene sequences taken from a well-characterized portion of the genome is used to identify a species

crop
: cultivated plant

heirloom seed
: seed from a plant that was grown historically, but has not been used in modern agriculture on a large scale

herbivory
: consumption of plants by insects and other animals

nectar
: liquid rich in sugars produced by flowers to attract animal pollinators

pollination

transfer of pollen from the anther to the stigma

Introduction
class="introduction"

The leaf chameleon (*Brookesia micra*) was discovered in northern Madagascar in 2012. At just over one inch long, it is the smallest known chameleon. (credit: modification of work by Frank Glaw, et al., PLOS)

While we can easily identify dogs, lizards, fish, spiders, and worms as animals, other animals, such as corals and sponges, might be easily mistaken as plants or some other form of life. Yet scientists have recognized a set of common characteristics shared by all animals, including sponges, jellyfish, sea urchins, and humans.

The kingdom Animalia is a group of multicellular Eukarya. Animal evolution began in the ocean over 600 million years ago, with tiny creatures that probably do not resemble any living organism today. Since then, animals have evolved into a highly diverse kingdom. Although over one million currently living species of animals have been identified, scientists are continually discovering more species. The number of described living animal species is estimated to be about 1.4 million,[footnote] and there may be as many as 6.8 million.
"Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

Understanding and classifying the variety of living species helps us to better understand how to conserve and benefit from this diversity. The animal classification system characterizes animals based on their anatomy, features of embryological development, and genetic makeup. Scientists are faced

with the task of classifying animals within a system of taxonomy that reflects their evolutionary history. Additionally, they must identify traits that are common to all animals as well as traits that can be used to distinguish among related groups of animals. However, animals vary in the complexity of their organization and exhibit a huge diversity of body forms, so the classification scheme is constantly changing as new information about species is learned.

Features of the Animal Kingdom

By the end of this section, you will be able to:

- List the features that distinguish the animal kingdom from other kingdoms
- Explain the processes of animal reproduction and embryonic development
- Describe the hierarchy of basic animal classification
- Compare and contrast the embryonic development of protostomes and deuterostomes

Even though members of the animal kingdom are incredibly diverse, animals share common features that distinguish them from organisms in other kingdoms. All animals are eukaryotic, multicellular organisms, and almost all animals have specialized tissues. Most animals are motile, at least during certain life stages. Animals require a source of food to grow and develop. All animals are heterotrophic, ingesting living or dead organic matter. This form of obtaining energy distinguishes them from autotrophic organisms, such as most plants, which make their own nutrients through photosynthesis and from fungi that digest their food externally. Animals may be carnivores, herbivores, omnivores, or parasites ([link]). Most animals reproduce sexually: The offspring pass through a series of developmental stages that establish a determined body plan, unlike plants, for example, in which the exact shape of the body is indeterminate. The **body plan** refers to the shape of an animal.

(a) (b)

All animals that derive energy from food are heterotrophs. The (a) black bear is an omnivore, eating both plants and animals. The (b) heartworm *Dirofilaria immitis* is a parasite that derives energy from its hosts. It spends its larval stage in mosquitos and its adult stage infesting the hearts of dogs and other mammals, as shown here. (credit a: modification of work by USDA Forest Service; credit b: modification of work by Clyde Robinson)

Complex Tissue Structure

A hallmark trait of animals is specialized structures that are differentiated to perform unique functions. As multicellular organisms, most animals develop specialized cells that group together into tissues with specialized functions. A tissue is a collection of similar cells that had a common embryonic origin. There are four main types of animal tissues: nervous, muscle, connective, and epithelial. Nervous tissue contains neurons, or nerve cells, which transmit nerve impulses. Muscle tissue contracts to cause all types of body movement from locomotion of the organism to movements within the body itself. Animals also have specialized connective tissues that provide many functions, including transport and structural support. Examples of connective tissues include blood and bone. Connective tissue is comprised of cells separated by extracellular material made of organic and inorganic materials, such as the protein and mineral deposits of bone. Epithelial tissue covers the internal and external surfaces of organs inside the animal body and the external surface of the body of the organism.

Note:
Concept in Action

View this video to watch a presentation by biologist E.O. Wilson on the importance of animal diversity.

Animal Reproduction and Development

Most animals have diploid body (somatic) cells and a small number of haploid reproductive (gamete) cells produced through meiosis. Some exceptions exist: For example, in bees, wasps, and ants, the male is haploid because it develops from an unfertilized egg. Most animals undergo sexual reproduction, while many also have mechanisms of asexual reproduction.

Sexual Reproduction and Embryonic Development

Almost all animal species are capable of reproducing sexually; for many, this is the only mode of reproduction possible. This distinguishes animals from fungi, protists, and bacteria, where asexual reproduction is common or exclusive. During sexual reproduction, the male and female gametes of a species combine in a process called fertilization. Typically, the small, motile male sperm travels to the much larger, sessile female egg. Sperm form is diverse and includes cells with flagella or amoeboid cells to facilitate motility. Fertilization and fusion of the gamete nuclei produce a zygote. Fertilization may be internal, especially in land animals, or external, as is common in many aquatic species.

After fertilization, a developmental sequence ensues as cells divide and differentiate. Many of the events in development are shared in groups of related animal species, and these events are one of the main ways scientists classify high-level groups of animals. During development, animal cells

specialize and form tissues, determining their future morphology and physiology. In many animals, such as mammals, the young resemble the adult. Other animals, such as some insects and amphibians, undergo complete metamorphosis in which individuals enter one or more larval stages. For these animals, the young and the adult have different diets and sometimes habitats. In other species, a process of incomplete metamorphosis occurs in which the young somewhat resemble the adults and go through a series of stages separated by molts (shedding of the skin) until they reach the final adult form.

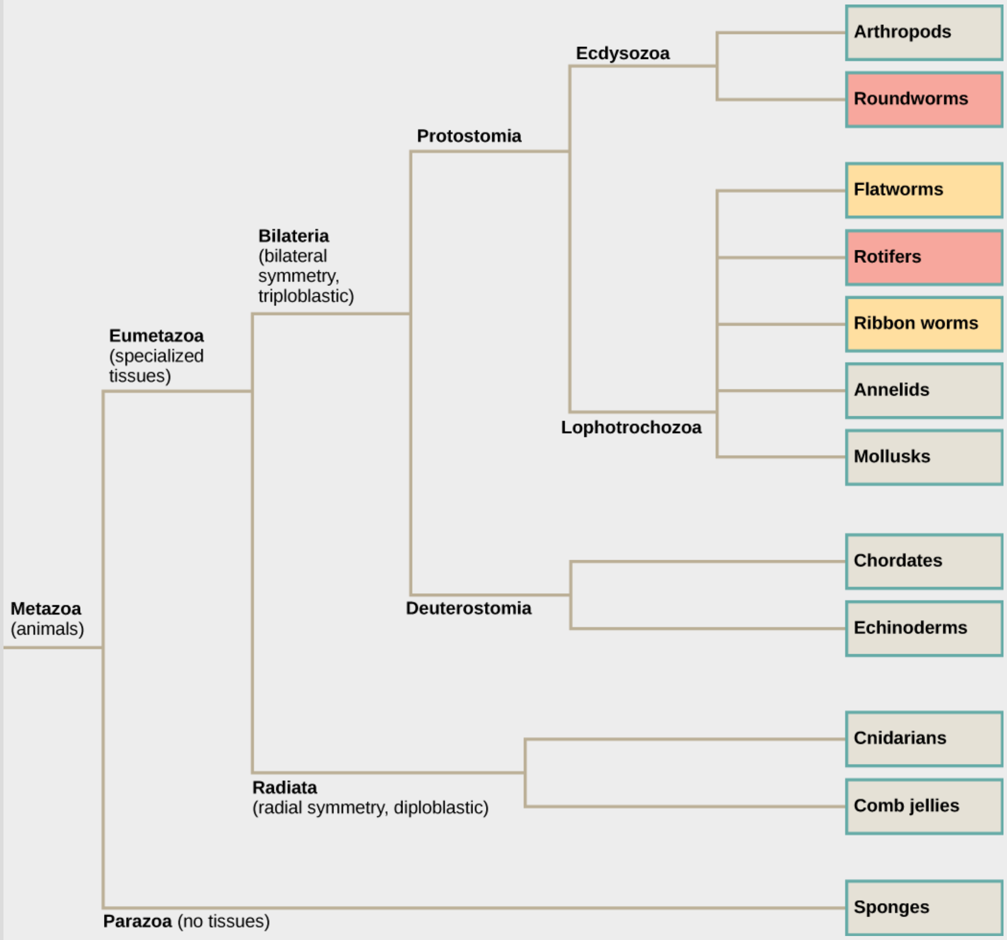
Asexual Reproduction

Asexual reproduction, unlike sexual reproduction, produces offspring genetically identical to each other and to the parent. A number of animal species—especially those without backbones, but even some fish, amphibians, and reptiles—are capable of asexual reproduction. Asexual reproduction, except for occasional identical twinning, is absent in birds and mammals. The most common forms of asexual reproduction for stationary aquatic animals include budding and fragmentation, in which part of a parent individual can separate and grow into a new individual. In contrast, a form of asexual reproduction found in certain invertebrates and rare vertebrates is called parthenogenesis (or "virgin beginning"), in which unfertilized eggs develop into new offspring.

Classification Features of Animals

Animals are classified according to morphological and developmental characteristics, such as a body plan. With the exception of sponges, the animal body plan is symmetrical. This means that their distribution of body parts is balanced along an axis. Additional characteristics that contribute to animal classification include the number of tissue layers formed during development, the presence or absence of an internal body cavity, and other features of embryological development.

Note:
Art Connection



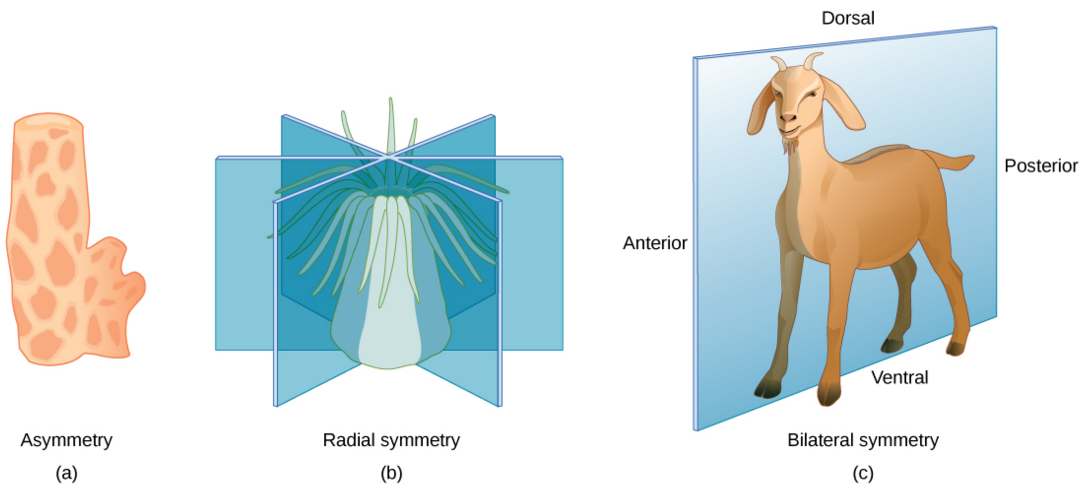
The phylogenetic tree of animals is based on morphological, fossil, and genetic evidence.

Which of the following statements is false?

a. Eumetazoa have specialized tissues and Parazoa do not.
b. Both acoelomates and pseudocoelomates have a body cavity.
c. Chordates are more closely related to echinoderms than to rotifers according to the figure.
d. Some animals have radial symmetry, and some animals have bilateral symmetry.

Body Symmetry

Animals may be asymmetrical, radial, or bilateral in form ([link]). **Asymmetrical** animals are animals with no pattern or symmetry; an example of an asymmetrical animal is a sponge ([link]**a**). An organism with **radial symmetry** ([link]**b**) has a longitudinal (up-and-down) orientation: Any plane cut along this up–down axis produces roughly mirror-image halves. An example of an organism with radial symmetry is a sea anemone.



Animals exhibit different types of body symmetry. The (a) sponge is asymmetrical and has no planes of symmetry, the (b) sea anemone has radial symmetry with multiple planes of symmetry, and the (c) goat has bilateral symmetry with one plane of symmetry.

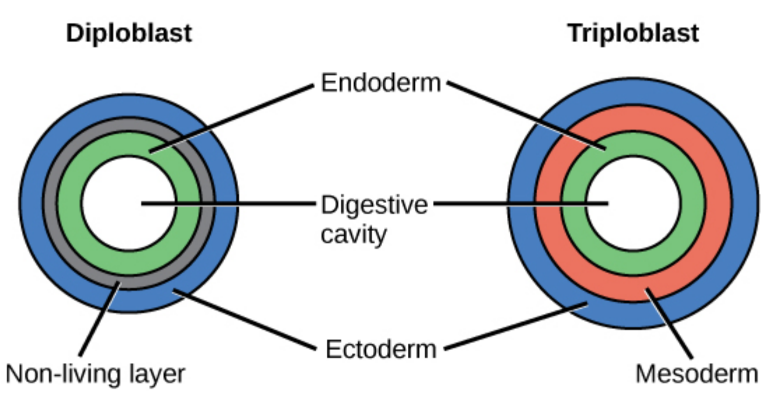
Bilateral symmetry is illustrated in [link]**c** using a goat. The goat also has upper and lower sides to it, but they are not symmetrical. A vertical plane cut from front to back separates the animal into roughly mirror-image right and left sides. Animals with bilateral symmetry also have a “head” and “tail” (anterior versus posterior) and a back and underside (dorsal versus ventral).

Note:
Concept in Action

Watch this video to see a quick sketch of the different types of body symmetry.

Layers of Tissues

Most animal species undergo a layering of early tissues during embryonic development. These layers are called **germ layers**. Each layer develops into a specific set of tissues and organs. Animals develop either two or three embryonic germs layers ([link]). The animals that display radial symmetry develop two germ layers, an inner layer (endoderm) and an outer layer (ectoderm). These animals are called **diploblasts**. Animals with bilateral symmetry develop three germ layers: an inner layer (endoderm), an outer layer (ectoderm), and a middle layer (mesoderm). Animals with three germ layers are called **triploblasts**.

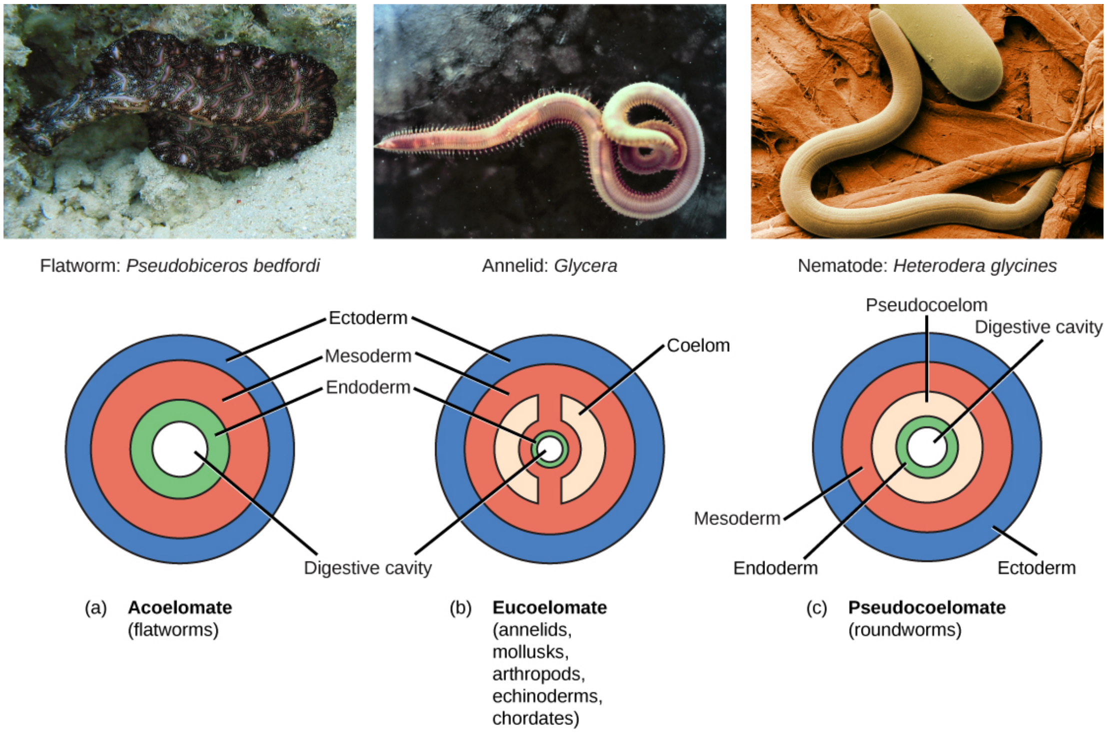


During embryogenesis, diploblasts

develop two embryonic germ layers: an ectoderm and an endoderm. Triploblasts develop a third layer—the mesoderm—between the endoderm and ectoderm.

Presence or Absence of a Coelom

Triploblasts may develop an internal body cavity derived from mesoderm, called a **coelom** (pr. see-LŌM). This epithelial-lined cavity is a space, usually filled with fluid, which lies between the digestive system and the body wall. It houses organs such as the kidneys and spleen, and contains the circulatory system. Triploblasts that do not develop a coelom are called **acoelomates**, and their mesoderm region is completely filled with tissue, although they have a gut cavity. Examples of acoelomates include the flatworms. Animals with a true coelom are called **eucoelomates** (or coelomates) ([link]). A true coelom arises entirely within the mesoderm germ layer. Animals such as earthworms, snails, insects, starfish, and vertebrates are all eucoelomates. A third group of triploblasts has a body cavity that is derived partly from mesoderm and partly from endoderm tissue. These animals are called **pseudocoelomates**. Roundworms are examples of pseudocoelomates. New data on the relationships of pseudocoelomates suggest that these phyla are not closely related and so the evolution of the pseudocoelom must have occurred more than once ([link]). True coelomates can be further characterized based on features of their early embryological development.

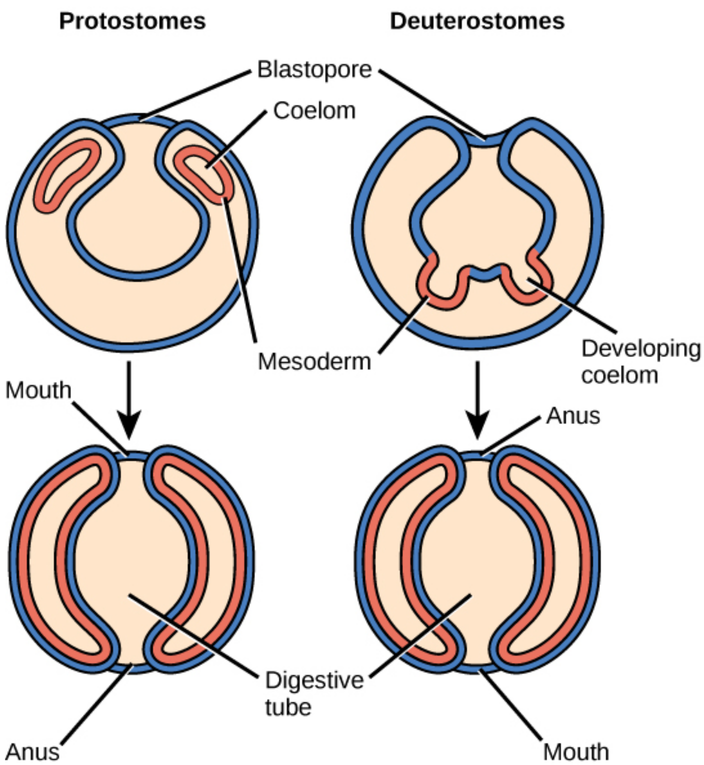


Triploblasts may be acoelomates, eucoelomates, or pseudocoelomates. Eucoelomates have a body cavity within the mesoderm, called a coelom, which is lined with mesoderm tissue. Pseudocoelomates have a similar body cavity, but it is lined with mesoderm and endoderm tissue. (credit a: modification of work by Jan Derk; credit b: modification of work by NOAA; credit c: modification of work by USDA, ARS)

Protostomes and Deuterostomes

Bilaterally symmetrical, triploblastic eucoelomates can be divided into two groups based on differences in their early embryonic development. **Protostomes** include phyla such as arthropods, mollusks, and annelids. **Deuterostomes** include the chordates and echinoderms. These two groups are named from which opening of the digestive cavity develops first: mouth

or anus. The word *protostome* comes from Greek words meaning “mouth first,” and *deuterostome* originates from words meaning “mouth second” (in this case, the anus develops first). This difference reflects the fate of a structure called the blastopore ([link]), which becomes the mouth in protostomes and the anus in deuterostomes. Other developmental characteristics differ between protostomes and deuterostomes, including the mode of formation of the coelom and the early cell division of the embryo.



Eucoelomates can be divided into two groups, protostomes and deuterostomes, based on their early embryonic development. Two of these differences include the origin of the mouth opening and the way in which the coelom is formed.

Section Summary

Animals constitute a diverse kingdom of organisms. Although animals range in complexity from simple sea sponges to human beings, most members share certain features. Animals are eukaryotic, multicellular, heterotrophic organisms that ingest their food and usually develop into motile creatures with a fixed body plan. Most members of the animal kingdom have differentiated tissues of four main classes—nervous, muscular, connective, and epithelial—that are specialized to perform different functions. Most animals reproduce sexually, leading to a developmental sequence that is relatively similar across the animal kingdom.

Organisms in the animal kingdom are classified based on their body morphology and development. True animals are divided into those with radial versus bilateral symmetry. Animals with three germ layers, called triploblasts, are further characterized by the presence or absence of an internal body cavity called a coelom. Animals with a body cavity may be either coelomates or pseudocoelomates, depending on which tissue gives rise to the coelom. Coelomates are further divided into two groups called protostomes and deuterostomes, based on a number of developmental characteristics.

Art Connection

Exercise:

Problem: [link] Which of the following statements is false?

a. Eumetazoa have specialized tissues and Parazoa do not.
b. Both acoelomates and pseudocoelomates have a body cavity.
c. Chordates are more closely related to echinoderms than to rotifers according to the figure.
d. Some animals have radial symmetry, and some animals have bilateral symmetry.

Solution:

[link] B

Review Questions

Exercise:

Problem:

Which of the following is not a feature common to *most* animals?

a. development into a fixed body plan
b. asexual reproduction
c. specialized tissues
d. heterotrophic nutrient sourcing

Solution:

B

Exercise:

Problem: Which of the following does not occur?

a. radially symmetrical diploblast
b. diploblastic eucoelomate
c. protostomic coelomate
d. bilaterally symmetrical deuterostome

Solution:

B

Free Response

Exercise:

Problem:

How are specialized tissues important for animal function and complexity?

Solution:

Specialized tissues allow more efficient functioning because differentiated tissue types can perform unique functions and work together in tandem to allow the animal to perform more functions. For example, specialized muscle tissue allows directed and efficient movement, and specialized nervous tissue allows for multiple sensory modalities as well as the ability to respond to various sensory information; these functions are not necessarily available to other non-animal organisms.

Exercise:

Problem:

Using the following terms, explain what classifications and groups humans fall into, from the most general to the most specific: symmetry, germ layers, coelom, embryological development.

Solution:

Humans have body plans that are bilaterally symmetrical and are characterized by the development of three germ layers, making them triploblasts. Humans have true coeloms, and are thus eucoelomates. Humans are deuterostomes.

Glossary

acoelomate
: without a body cavity

asymmetrical

having no plane of symmetry

bilateral symmetry
: a type of symmetry in which there is only one plane of symmetry that creates two mirror-image sides

body plan
: the shape and symmetry of an organism

coelom
: a lined body cavity derived from mesodermal embryonic tissue

deuterostome
: describing an animal in which the blastopore develops into the anus, with the second opening developing into the mouth

diploblast
: an animal that develops from two embryonic germ layers

eucoelomate
: describing animals with a body cavity completely lined with mesodermal tissue

germ layer
: a collection of cells formed during embryogenesis that will give rise to future body tissues

protostome
: describing an animal in which the mouth develops first during embryogenesis and a second opening developing into the anus

pseudocoelomate
: an animal with a coelom that is not completely lined with tissues derived from the mesoderm as in eucoelomate animals

radial symmetry
: a type of symmetry with multiple planes of symmetry all cross at an axis through the center of the organism

triploblast

an animal that develops from three germ layers

Sponges and Cnidarians
By the end of this section, you will be able to:

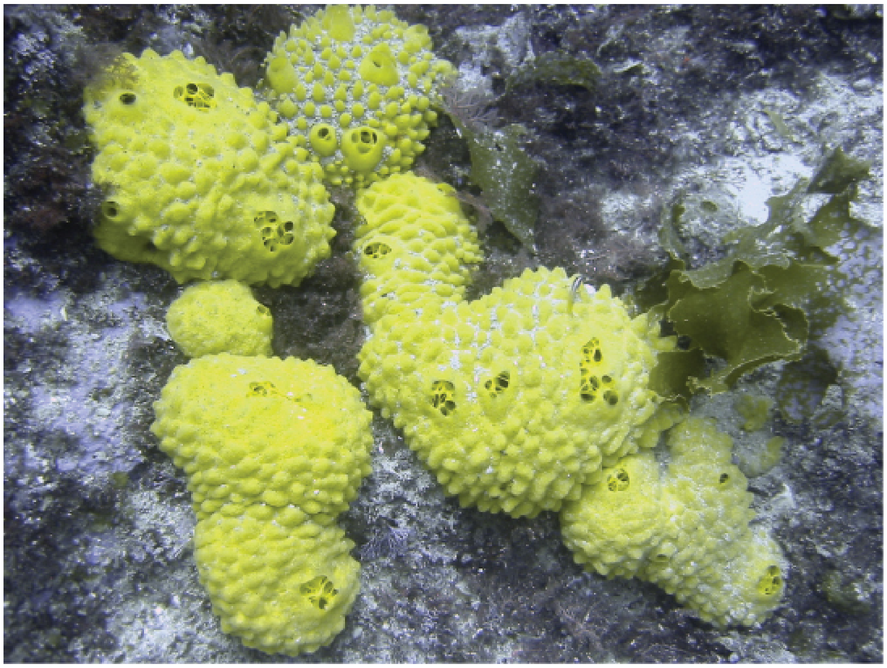
- Describe the organizational features of the simplest animals
- Describe the organizational features of cnidarians

The kingdom of animals is informally divided into invertebrate animals, those without a backbone, and vertebrate animals, those with a backbone. Although in general we are most familiar with vertebrate animals, the vast majority of animal species, about 95 percent, are invertebrates. Invertebrates include a huge diversity of animals, millions of species in about 32 phyla, which we can just begin to touch on here.

The sponges and the cnidarians represent the simplest of animals. Sponges appear to represent an early stage of multicellularity in the animal clade. Although they have specialized cells for particular functions, they lack true tissues in which specialized cells are organized into functional groups. Sponges are similar to what might have been the ancestor of animals: colonial, flagellated protists. The cnidarians, or the jellyfish and their kin, are the simplest animal group that displays true tissues, although they possess only two tissue layers.

Sponges

Animals in subkingdom Parazoa represent the simplest animals and include the sponges, or phylum **Porifera** ([link]). All sponges are aquatic and the majority of species are marine. Sponges live in intimate contact with water, which plays a role in their feeding, gas exchange, and excretion. Much of the body structure of the sponge is dedicated to moving water through the body so it can filter out food, absorb dissolved oxygen, and eliminate wastes.

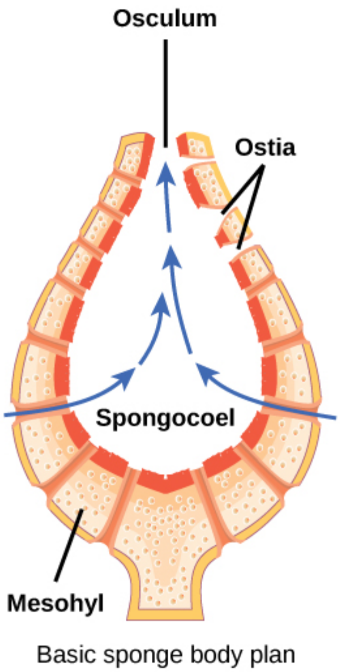


Sponges are members of the phylum Porifera, which contains the simplest animals. (credit: Andrew Turner)

The body of the simplest sponges takes the shape of a cylinder with a large central cavity, the **spongocoel**. Water enters the spongocoel from numerous pores in the body wall. Water flows out through a large opening called the **osculum** ([link]). However, sponges exhibit a diversity of body forms, which vary in the size and branching of the spongocoel, the number of osculi, and where the cells that filter food from the water are located.

Sponges consist of an outer layer of flattened cells and an inner layer of cells called choanocytes separated by a jelly-like substance called **mesohyl**. The mesohyl contains embedded amoeboid cells that secrete tiny needles called **spicules** or protein fibers that help give the sponge its structural strength. The cell body of the **choanocyte** is embedded in mesohyl but protruding into the spongocoel is a mesh-like collar surrounding a single flagellum. The beating of flagella from all choanocytes moves water through the sponge. Food particles are trapped in mucus produced by the sieve-like collar of the choanocytes and are ingested by phagocytosis. This

process is called **intracellular digestion**. **Amoebocytes** take up nutrients repackaged in food vacuoles of the choanocytes and deliver them to other cells within the sponge.



The sponge's basic body plan is shown.

Physiological Processes in Sponges

Despite their lack of complexity, sponges are clearly successful organisms, having persisted on Earth for more than half a billion years. Lacking a true digestive system, sponges depend on the intracellular digestive processes of their choanocytes for their energy intake. The limit of this type of digestion is that food particles must be smaller than individual cells. Gas exchange, circulation, and excretion occur by diffusion between cells and the water.

Sponges reproduce both sexually and asexually. Asexual reproduction is either by **fragmentation** (in which a piece of the sponge breaks off and develops into a new individual), or **budding** (an outgrowth from the parent that eventually detaches). A type of asexual reproduction found only in freshwater sponges occurs through the formation of **gemmules**, clusters of cells surrounded by a tough outer layer. Gemmules survive hostile environments and can attach to a substrate and grow into a new sponge.

Sponges are **monoecious** (or hermaphroditic), meaning one individual can produce both eggs and sperm. Sponges may be sequentially hermaphroditic, producing eggs first and sperm later. Eggs arise from amoebocytes and are retained within the spongocoel, whereas sperm arise from choanocytes and are ejected through the osculum. Sperm carried by water currents fertilize the eggs of other sponges. Early larval development occurs within the sponge, and free-swimming larvae are then released through the osculum. This is the only time that sponges exhibit mobility. Sponges are sessile as adults and spend their lives attached to a fixed substrate.

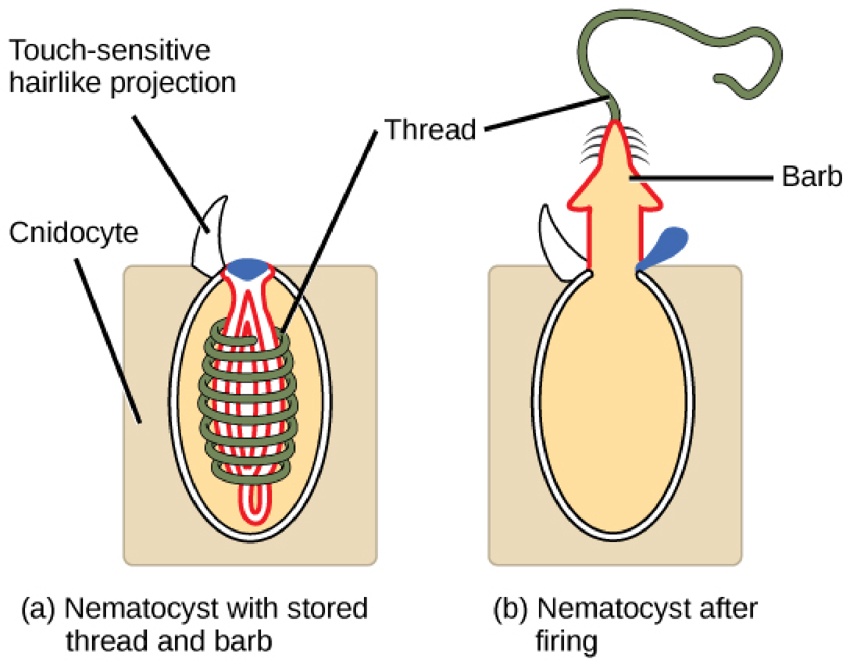
Note:
Concept in Action

Watch this video that demonstrates the feeding of sponges.

Cnidarians

The phylum **Cnidaria** includes animals that show radial or biradial symmetry and are diploblastic. Nearly all (about 99 percent) cnidarians are marine species. Cnidarians have specialized cells known as **cnidocytes**

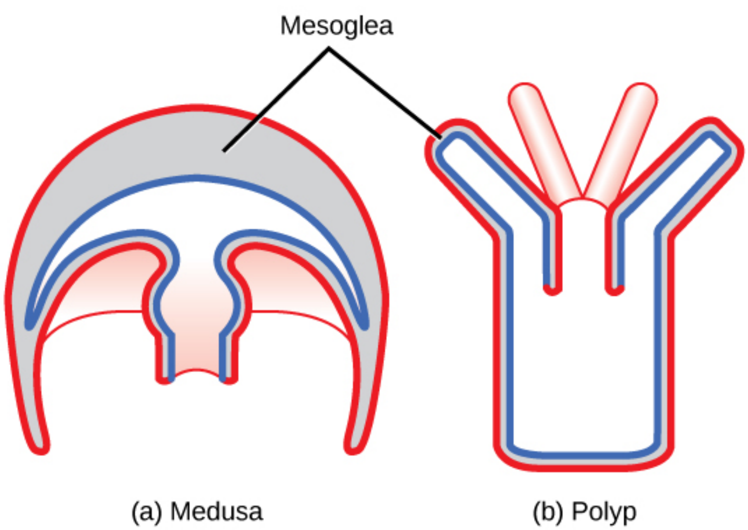
(“stinging cells”) containing organelles called **nematocysts**. These cells are concentrated around the mouth and tentacles of the animal and can immobilize prey with toxins. Nematocysts contain coiled threads that may bear barbs. The outer wall of the cell has a hairlike projection that is sensitive to touch. When touched, the cells fire the toxin-containing coiled threads that can penetrate and stun the predator or prey (see [link]).



Animals from the phylum Cnidaria have stinging cells called cnidocytes. Cnidocytes contain large organelles called (a) nematocysts that store a coiled thread and barb. When hairlike projections on the cell surface are touched, (b) the thread, barb, and a toxin are fired from the organelle.

Cnidarians display two distinct body plans: **polyp** or “stalk” and **medusa** or “bell” ([link]). Examples of the polyp form are freshwater species of the genus *Hydra*; perhaps the best-known medusoid animals are the jellies

(jellyfish). Polyps are sessile as adults, with a single opening to the digestive system (the mouth) facing up with tentacles surrounding it. Medusae are motile, with the mouth and tentacles hanging from the bell-shaped body. In other cnidarians, both a polyp and medusa form exist, and the life cycle alternates between these forms.



Cnidarians have two distinct body plans, the (a) medusa and the (b) polyp. All cnidarians have two tissue layers, with a jelly-like mesoglea between them.

Physiological Processes of Cnidarians

All cnidarians have two tissue layers. The outer layer is called the **epidermis**, whereas the inner layer is called the **gastrodermis** and lines the digestive cavity. Between these two layers is a non-living, jelly-like **mesoglea**. There are differentiated cell types in each tissue layer, such as nerve cells, enzyme-secreting cells, and nutrient-absorbing cells, as well as

intercellular connections between the cells. However, organs and organ systems are not present in this phylum.

The nervous system is primitive, with nerve cells scattered across the body in a network. The function of the nerve cells is to carry signals from sensory cells and to contractile cells. Groups of cells in the nerve net form nerve cords that may be essential for more rapid transmission. Cnidarians perform **extracellular digestion**, with digestion completed by intracellular digestive processes. Food is taken into the **gastrovascular cavity**, enzymes are secreted into the cavity, and the cells lining the cavity absorb the nutrient products of the extracellular digestive process. The gastrovascular cavity has only one opening that serves as both a mouth and an anus (an incomplete digestive system). Like the sponges, Cnidarian cells exchange oxygen, carbon dioxide, and nitrogenous wastes by diffusion between cells in the epidermis and gastrodermis with water.

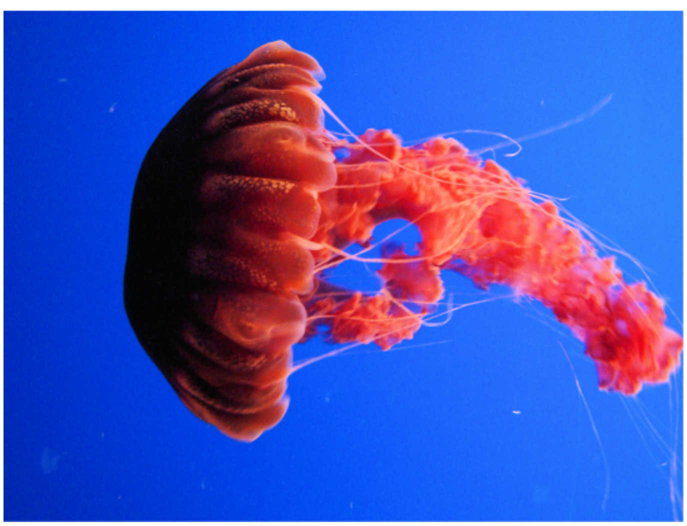
Cnidarian Diversity

The phylum Cnidaria contains about 10,000 described species divided into four classes: Anthozoa, Scyphozoa, Cubozoa, and Hydrozoa.

The class Anthozoa includes all cnidarians that exhibit a sessile polyp body plan only; in other words, there is no medusa stage within their life cycle. Examples include sea anemones, sea pens, and corals, with an estimated number of 6,100 described species. Sea anemones are usually brightly colored and can attain a size of 1.8 to 10 cm in diameter. These animals are usually cylindrical in shape and are attached to a substrate. A mouth opening is surrounded by tentacles bearing cnidocytes ([link]).

Sea anemones are cnidarians of class Anthozoa. (credit: "Dancing With Ghosts"/Flickr)

Scyphozoans include all the jellies and are motile and exclusively marine with about 200 described species. The medusa is the dominant stage in the life cycle, although there is also a polyp stage. Species range from 2 cm in length to the largest scyphozoan species, *Cyanea capillata*, at 2 m across. Jellies display a characteristic bell-like body shape ([link]).



Scyphozoans include the jellies. (credit: "Jimg944"/Flickr)

Note:
Concept in Action

Use this video to identify the life cycle stages of jellies.

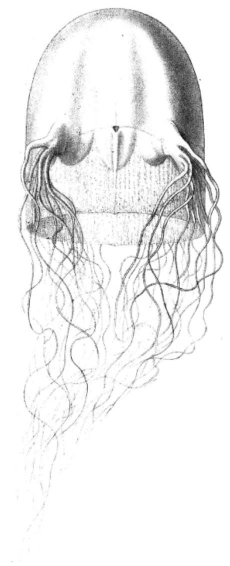
The class Cubozoa includes jellies that are square in cross-section and so are known as “box jellyfish.” These species may achieve sizes of 15–25 cm. Cubozoans are anatomically similar to the jellyfish. A prominent difference between the two classes is the arrangement of tentacles. Cubozoans have muscular pads called pedalia at the corners of the square bell canopy, with one or more tentacles attached to each pedalium. In some cases, the digestive system may extend into the pedalia. Cubozoans typically exist in a polyp form that develops from a larva. The polyps may bud to form more polyps and then transform into the medusoid forms.

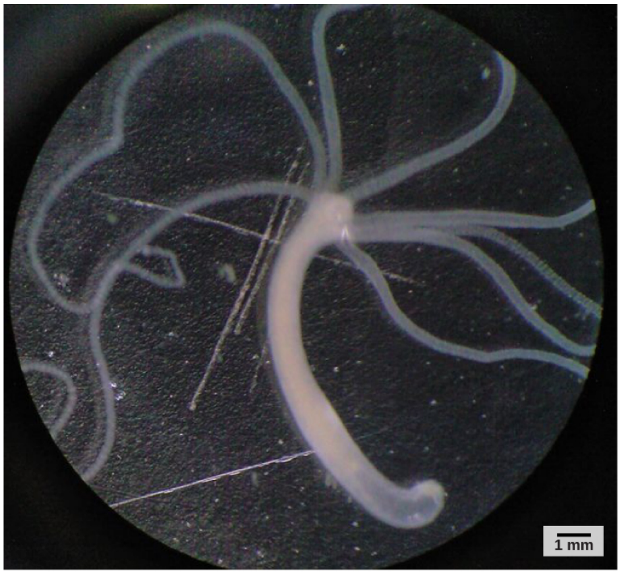
Note:
Concept in Action

Watch this video to learn more about the deadly toxins of the box jellyfish.

Hydrozoa includes nearly 3,500 species,[footnote] most of which are marine. Most species in this class have both polyp and medusa forms in their life cycle. Many hydrozoans form colonies composed of branches of specialized polyps that share a gastrovascular cavity. Colonies may also be free-floating and contain both medusa and polyp individuals in the colony, as in the Portuguese Man O'War (*Physalia*) or By-the-Wind Sailor (*Velella*). Other species are solitary polyps or solitary medusae. The characteristic shared by all of these species is that their gonads are derived from epidermal tissue, whereas in all other cnidarians, they are derived from gastrodermal tissue ([link]**ab**).

"The Hydrozoa Directory," Peter Schuchert, Muséum Genève, last updated November 2012, http://www.ville-ge.ch/mhng/hydrozoa/hydrozoa-directory.htm.


(a)



(b)

A (a) box jelly is an example from class Cubozoa. The (b) hydra is from class Hydrozoa. (credit b: scale-bar data from Matt Russell)

Section Summary

Animals included in phylum Porifera are parazoans and do not possess true tissues. These organisms show a simple organization. Sponges have multiple cell types that are geared toward executing various metabolic functions.

Cnidarians have outer and inner tissue layers sandwiching a noncellular mesoglea. Cnidarians possess a well-formed digestive system and carry out extracellular digestion. The cnidocyte is a specialized cell for delivering toxins to prey and predators. Cnidarians have separate sexes. They have a life cycle that involves morphologically distinct forms—medusoid and polypoid—at various stages in their life cycle.

Review Questions

Exercise:

Problem:

The large central opening in the poriferan body is called the _____.

a. emmule
b. picule
c. stia
d. osculum

Solution:

D

Exercise:

Problem: Cnidocytes are found in _____.

a. phylum Porifera
b. phylum Nemertea
c. phylum Nematoda
d. phylum Cnidaria

Solution:

D

Exercise:

Problem: Cubozoans are ________.

a. polyps
b. medusoids
c. polymorphs
d. sponges

Solution:

B

Free Response

Exercise:

Problem:

Describe the feeding mechanism of sponges and identify how it is different from other animals.

Solution:

The sponges draw water carrying food particles into the spongocoel using the beating of flagella in the choanocytes. The food particles are caught by the collar of the choanocyte and brought into the cell by phagocytosis. Digestion of the food particle takes place inside the cell. The difference between this and the mechanisms of other animals is that digestion takes place within cells rather than outside of cells. It means that the organism can feed only on particles smaller than the cells themselves.

Exercise:

Problem:

Compare the structural differences between Porifera and Cnidaria.

Solution:

Poriferans do not possess true tissues, whereas cnidarians do have tissues. Because of this difference, poriferans do not have a nerve net or muscle cells for locomotion, which cnidarians have.

Glossary

amoebocyte

an amoeba-like cell of sponges whose functions include distribution of nutrients to other cells in the sponge

budding

a form of asexual reproduction that occurs through the growth of a new organism as a branch on an adult organism that breaks off and becomes independent; found in plants, sponges, cnidarians, and some other invertebrates

choanocyte

a cell type unique to sponges with a flagellum surrounded by a collar used to maintain water flow through the sponge, and capture and digest food particles

Cnidaria

a phylum of animals that are diploblastic and have radial symmetry and stinging cells

cnidocyte

a specialized stinging cell found in Cnidaria

epidermis

the layer of cells that lines the outer surface of an animal

extracellular digestion

a form of digestion, the breakdown of food, which occurs outside of cells with the aid of enzymes released by cells

fragmentation

a form of asexual reproduction in which a portion of the body of an organism breaks off and develops into a living independent organism; found in plants, sponges, and some other invertebrates

gastrodermis

the layer of cells that lines the gastrovascular cavity of cnidarians

gastrovascular cavity

the central cavity bounded by the gastrodermis in cnidarians

gemmule
a structure produced by asexual reproduction in freshwater sponges that is able to survive harsh conditions

intracellular digestion
the digestion of matter brought into a cell by phagocytosis

medusa
a free-floating cnidarian body plan with a mouth on the underside and tentacles hanging down from a bell

mesoglea
the non-living, gel-like matrix present in between ectoderm and endoderm in cnidarians

mesohyl
the collagen-like gel containing suspended cells that perform various functions in sponges

monoecious
having both sexes in one body, hermaphroditic

nematocyst
the harpoon-like organelle within a cnidocyte with a pointed projectile and poison to stun and entangle prey

osculum
the large opening in a sponge body through which water leaves

polyp
the stalk-like, sessile life form of a cnidarians with mouth and tentacles facing upward, usually sessile but may be able to glide along a surface

Porifera
a phylum of animals with no true tissues, but a porous body with a rudimentary endoskeleton

spicule
: a short sliver or spike-like structure, in sponges, they are formed of silicon dioxide, calcium carbonate, or protein, and are found in the mesohyl

spongocoel
: the central cavity within the body of some sponges

Flatworms, Nematodes, and Arthropods
By the end of this section, you will be able to:

- Describe the structure and systems of flatworms
- Describe the structural organization of nematodes
- Compare the internal systems and the appendage specialization of arthropods

The animal phyla of this and subsequent modules are triploblastic and have an embryonic mesoderm sandwiched between the ectoderm and endoderm. These phyla are also bilaterally symmetrical, meaning that a longitudinal section will divide them into right and left sides that are mirror images of each other. Associated with bilateralism is the beginning of cephalization, the evolution of a concentration of nervous tissues and sensory organs in the head of the organism, which is where the organism first encounters its environment.

The flatworms are acoelomate organisms that include free-living and parasitic forms. The nematodes, or roundworms, possess a pseudocoelom and consist of both free-living and parasitic forms. Finally, the arthropods, one of the most successful taxonomic groups on the planet, are coelomate organisms with a hard exoskeleton and jointed appendages. The nematodes and the arthropods belong to a clade with a common ancestor, called Ecdysozoa. The name comes from the word *ecdysis*, which refers to the periodic shedding, or molting, of the exoskeleton. The ecdysozoan phyla have a hard cuticle covering their bodies that must be periodically shed and replaced for them to increase in size.

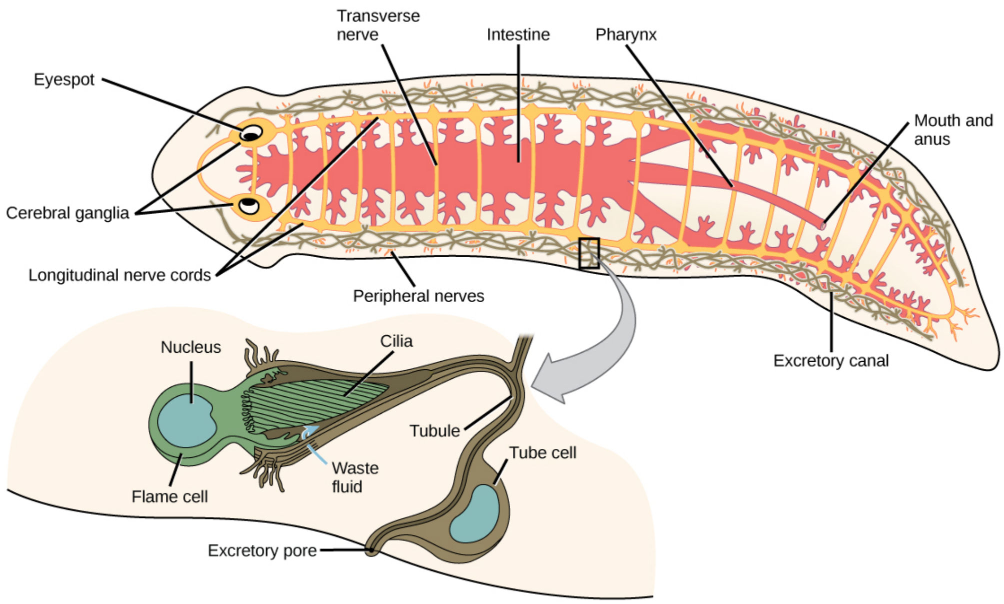
Flatworms

The relationships among flatworms, or phylum Platyhelminthes, is being revised and the description here will follow the traditional groupings. Most flatworms are parasitic, including important parasites of humans. Flatworms have three embryonic germ layers that give rise to surfaces covering tissues, internal tissues, and the lining of the digestive system. The epidermal tissue is a single layer of cells or a layer of fused cells covering a layer of circular muscle above a layer of longitudinal muscle. The

mesodermal tissues include support cells and secretory cells that secrete mucus and other materials to the surface. The flatworms are acoelomate, so their bodies contain no cavities or spaces between the outer surface and the inner digestive tract.

Physiological Processes of Flatworms

Free-living species of flatworms are predators or scavengers, whereas parasitic forms feed from the tissues of their hosts. Most flatworms have an incomplete digestive system with an opening, the “mouth,” that is also used to expel digestive system wastes. Some species also have an anal opening. The gut may be a simple sac or highly branched. Digestion is extracellular, with enzymes secreted into the space by cells lining the tract, and digested materials taken into the same cells by phagocytosis. One group, the cestodes, does not have a digestive system, because their parasitic lifestyle and the environment in which they live (suspended within the digestive cavity of their host) allows them to absorb nutrients directly across their body wall. Flatworms have an excretory system with a network of tubules throughout the body that open to the environment and nearby flame cells, whose cilia beat to direct waste fluids concentrated in the tubules out of the body. The system is responsible for regulation of dissolved salts and excretion of nitrogenous wastes. The nervous system consists of a pair of nerve cords running the length of the body with connections between them and a large ganglion or concentration of nerve cells at the anterior end of the worm; here, there may also be a concentration of photosensory and chemosensory cells ([link]).



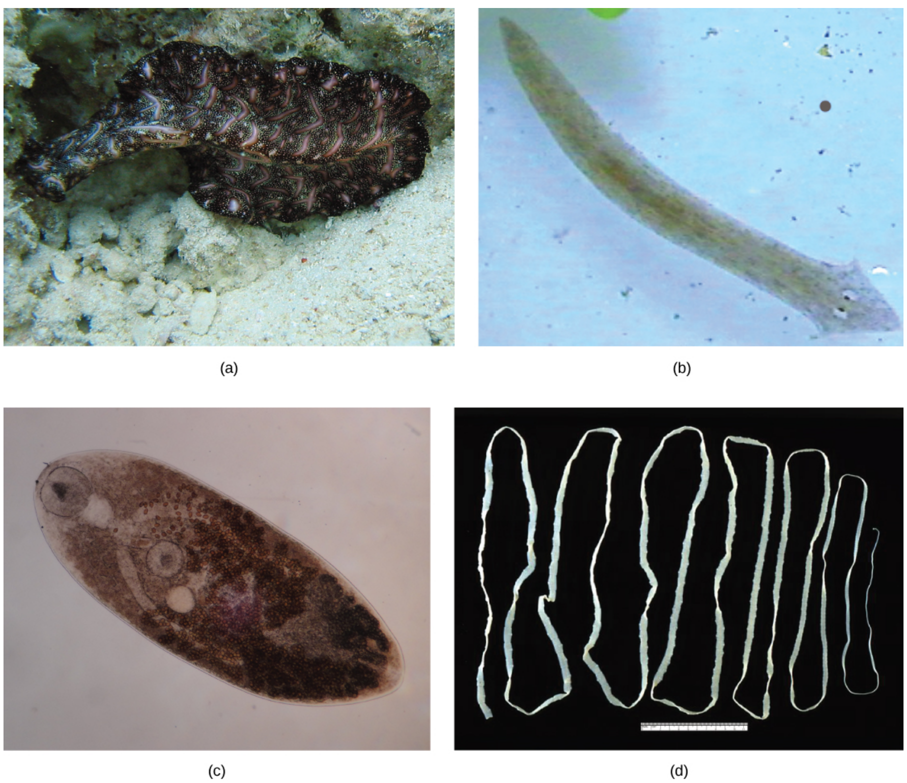
This planarian is a free-living flatworm that has an incomplete digestive system, an excretory system with a network of tubules throughout the body, and a nervous system made up of nerve cords running the length of the body with a concentration of nerves and photosensory and chemosensory cells at the anterior end.

Since there is no circulatory or respiratory system, gas and nutrient exchange is dependent on diffusion and intercellular junctions. This necessarily limits the thickness of the body in these organisms, constraining them to be “flat” worms. Most flatworm species are monoecious (hermaphroditic, possessing both sets of sex organs), and fertilization is typically internal. Asexual reproduction is common in some groups in which an entire organism can be regenerated from just a part of itself.

Diversity of Flatworms

Flatworms are traditionally divided into four classes: Turbellaria, Monogenea, Trematoda, and Cestoda ([link]). The turbellarians include

mainly free-living marine species, although some species live in freshwater or moist terrestrial environments. The simple planarians found in freshwater ponds and aquaria are examples. The epidermal layer of the underside of turbellarians is ciliated, and this helps them move. Some turbellarians are capable of remarkable feats of regeneration in which they may regrow the body, even from a small fragment.



Phylum Platyhelminthes is divided into four classes: (a) Bedford's Flatworm (*Pseudobiceros bedfordi*) and the (b) planarian belong to class Turbellaria; (c) the Trematoda class includes about 20,000 species, most of which are parasitic; (d) class Cestoda includes tapeworms such as this *Taenia saginata*; and the parasitic class Monogenea (not shown). (credit a: modification of work by Jan Derk; credit c: modification of work by "Sahaquiel9102"/Wikimedia Commons; credit d: modification of work by CDC)

The monogeneans are external parasites mostly of fish with life cycles consisting of a free-swimming larva that attaches to a fish to begin transformation to the parasitic adult form. They have only one host during their life, typically of just one species. The worms may produce enzymes that digest the host tissues or graze on surface mucus and skin particles. Most monogeneans are hermaphroditic, but the sperm develop first, and it is typical for them to mate between individuals and not to self-fertilize.

The trematodes, or flukes, are internal parasites of mollusks and many other groups, including humans. Trematodes have complex life cycles that involve a primary host in which sexual reproduction occurs and one or more secondary hosts in which asexual reproduction occurs. The primary host is almost always a mollusk. Trematodes are responsible for serious human diseases including schistosomiasis, caused by a blood fluke (*Schistosoma*). The disease infects an estimated 200 million people in the tropics and leads to organ damage and chronic symptoms including fatigue. Infection occurs when a human enters the water, and a larva, released from the primary snail host, locates and penetrates the skin. The parasite infects various organs in the body and feeds on red blood cells before reproducing. Many of the eggs are released in feces and find their way into a waterway where they are able to reinfect the primary snail host.

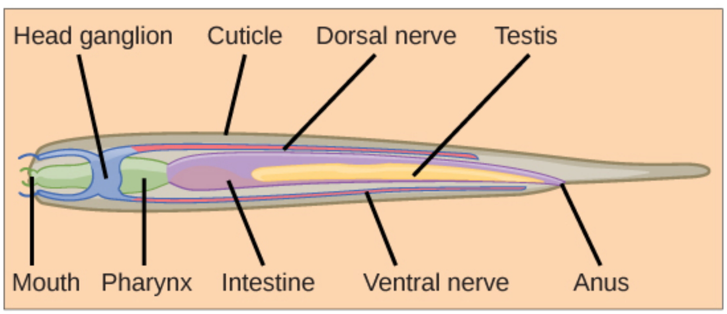
The cestodes, or tapeworms, are also internal parasites, mainly of vertebrates. Tapeworms live in the intestinal tract of the primary host and remain fixed using a sucker on the anterior end, or scolex, of the tapeworm body. The remaining body of the tapeworm is made up of a long series of units called proglottids, each of which may contain an excretory system with flame cells, but will contain reproductive structures, both male and female. Tapeworms do not have a digestive system, they absorb nutrients from the food matter passing them in the host’s intestine. Proglottids are produced at the scolex and are pushed to the end of the tapeworm as new proglottids form, at which point, they are “mature” and all structures except fertilized eggs have degenerated. Most reproduction occurs by cross-fertilization. The proglottid detaches and is released in the feces of the host.

The fertilized eggs are eaten by an intermediate host. The juvenile worms emerge and infect the intermediate host, taking up residence, usually in muscle tissue. When the muscle tissue is eaten by the primary host, the cycle is completed. There are several tapeworm parasites of humans that are acquired by eating uncooked or poorly cooked pork, beef, and fish.

Nematodes

The phylum **Nematoda**, or roundworms, includes more than 28,000 species with an estimated 16,000 parasitic species. The name Nematoda is derived from the Greek word “nemos,” which means “thread.” Nematodes are present in all habitats and are extremely common, although they are usually not visible ([link]).

(a)



(b)

(a) An scanning electron micrograph of the nematode *Heterodera glycines* and (b) a schematic representation of the anatomy of a nematode are shown. (credit a: modification of work by USDA, ARS; scale-bar data from Matt Russell)

Most nematodes look similar to each other: slender tubes, tapered at each end ([link]). Nematodes are pseudocoelomates and have a **complete digestive system** with a distinct mouth and anus.

The nematode body is encased in a cuticle, a flexible but tough exoskeleton, or external skeleton, which offers protection and support. The cuticle contains a carbohydrate-protein polymer called **chitin**. The cuticle also lines the pharynx and rectum. Although the exoskeleton provides protection, it restricts growth, and therefore must be continually shed and replaced as the animal increases in size.

A nematode's mouth opens at the anterior end with three or six lips and, in some species, teeth in the form of cuticular extensions. There may also be a sharp stylet that can protrude from the mouth to stab prey or pierce plant or animal cells. The mouth leads to a muscular pharynx and intestine, leading to the rectum and anal opening at the posterior end.

Physiological Processes of Nematodes

In nematodes, the excretory system is not specialized. Nitrogenous wastes are removed by diffusion. In marine nematodes, regulation of water and salt is achieved by specialized glands that remove unwanted ions while maintaining internal body fluid concentrations.

Most nematodes have four nerve cords that run along the length of the body on the top, bottom, and sides. The nerve cords fuse in a ring around the pharynx, to form a head ganglion or "brain" of the worm, as well as at the posterior end to form the tail ganglion. Beneath the epidermis lies a layer of longitudinal muscles that permits only side-to-side, wave-like undulation of the body.

Note:
Concept in Action

View this video to see nematodes move about and feed on bacteria.

Nematodes employ a diversity of sexual reproductive strategies depending on the species; they may be monoecious, **dioecious** (separate sexes), or may reproduce asexually by parthenogenesis. *Caenorhabditis elegans* is nearly unique among animals in having both self-fertilizing hermaphrodites and a male sex that can mate with the hermaphrodite.

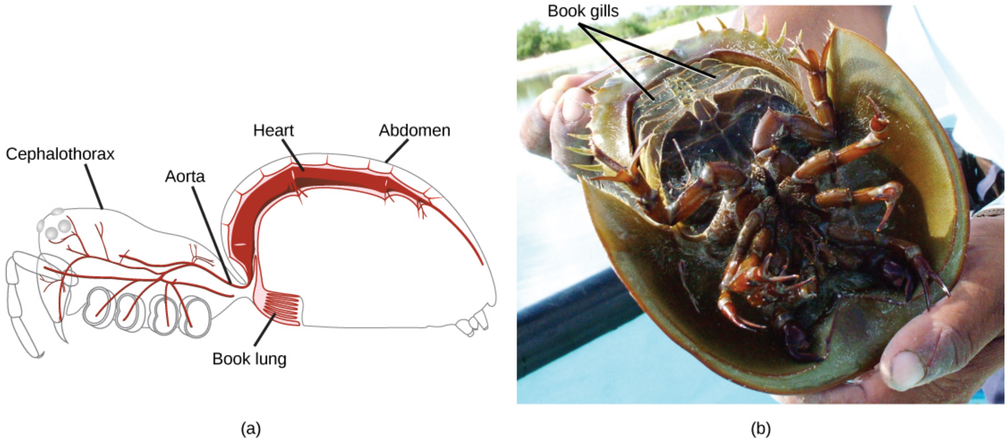
Arthropoda

The name “arthropoda” means “jointed legs,” which aptly describes each of the enormous number of species belonging to this phylum. **Arthropoda** dominate the animal kingdom with an estimated 85 percent of known species, with many still undiscovered or undescribed. The principal characteristics of all the animals in this phylum are functional segmentation of the body and the presence of jointed appendages ([link]). As members of Ecdysozoa, arthropods also have an exoskeleton made principally of chitin. Arthropoda is the largest phylum in the animal world in terms of numbers of species, and insects form the single largest group within this phylum. Arthropods are true coelomate animals and exhibit prostostomic development.

Trilobites, like the one in this fossil, are an extinct group of arthropods. (credit: Kevin Walsh)

Physiological Processes of Arthropods

A unique feature of arthropods is the presence of a segmented body with fusion of certain sets of segments to give rise to functional segments. Fused segments may form a head, thorax, and abdomen, or a cephalothorax and abdomen, or a head and trunk. The coelom takes the form of a **hemocoel** (or blood cavity). The open circulatory system, in which blood bathes the internal organs rather than circulating in vessels, is regulated by a two-chambered heart. Respiratory systems vary, depending on the group of arthropod: Insects and myriapods use a series of tubes (**tracheae**) that branch throughout the body, open to the outside through openings called **spiracles**, and perform gas exchange directly between the cells and air in the tracheae. Aquatic crustaceans use gills, arachnids employ “book lungs,” and aquatic chelicerates use “book gills.” The book lungs of arachnids are internal stacks of alternating air pockets and hemocoel tissue shaped like the pages of a book. The book gills of crustaceans are external structures similar to book lungs with stacks of leaf-like structures that exchange gases with the surrounding water ([link]).

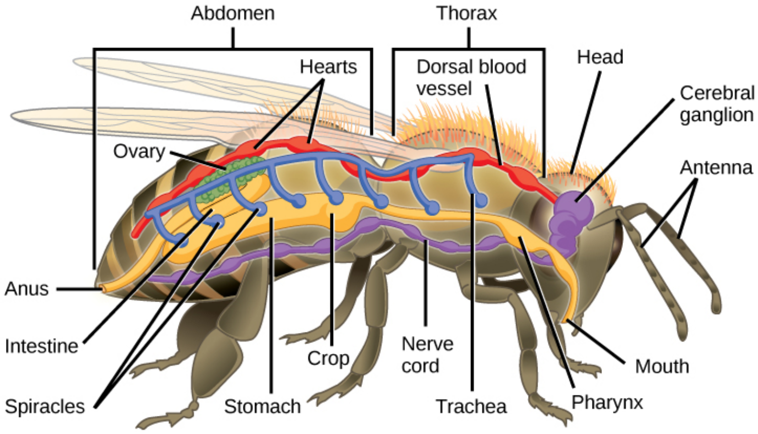


The book lungs of (a) arachnids are made up of alternating air pockets and hemocoel tissue shaped like a stack of books. The book gills of (b) crustaceans are similar to book lungs but are external so that gas exchange can occur with the surrounding water. (credit a: modification of work by Ryan Wilson based on original work by John Henry Comstock; credit b: modification of work by Angel Schatz)

Arthropod Diversity

Phylum Arthropoda includes animals that have been successful in colonizing terrestrial, aquatic, and aerial habitats. The phylum is further classified into five subphyla: Trilobitomorpha (trilobites), Hexapoda (insects and relatives), Myriapoda (millipedes, centipedes, and relatives), Crustacea (crabs, lobsters, crayfish, isopods, barnacles, and some zooplankton), and Chelicerata (horseshoe crabs, arachnids, scorpions, and daddy longlegs). Trilobites are an extinct group of arthropods found from the Cambrian period (540–490 million years ago) until they became extinct in the Permian (300–251 million years ago) that are probably most closely related to the Chelicerata. The 17,000 described species have been identified from fossils ([link]).

The Hexapoda have six legs (three pairs) as their name suggests. Hexapod segments are fused into a head, thorax, and abdomen ([link]). The thorax bears the wings and three pairs of legs. The insects we encounter on a daily basis—such as ants, cockroaches, butterflies, and bees—are examples of Hexapoda.



In this basic anatomy of a hexapod, note that insects have a developed digestive system (yellow), a respiratory system (blue), a circulatory system (red), and a nervous system (purple).

Subphylum Myriapoda includes arthropods with legs that may vary in number from 10 to 750. This subphylum includes 13,000 species; the most commonly found examples are millipedes and centipedes. All myriapods are terrestrial animals and prefer a humid environment ([link]).

(a) (b)

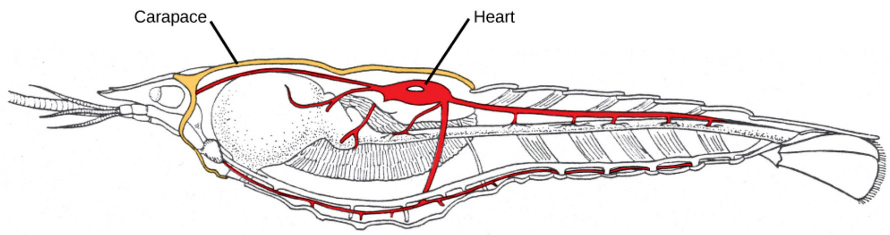
(a) The centipede *Scutigera coleoptrata* has up to 15 pairs of legs. (b) This North American millipede (*Narceus americanus)* bears many legs, although not one thousand, as its name might suggest. (credit a: modification of work by Bruce Marlin; credit b: modification of work by Cory Zanker)

Crustaceans, such as shrimp, lobsters, crabs, and crayfish, are the dominant aquatic arthropods. A few crustaceans are terrestrial species like the pill bugs or sow bugs. The number of described crustacean species stands at about 47,000.[footnote]

“Number of Living Species in Australia and the World,” A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

Although the basic body plan in crustaceans is similar to the Hexapoda—head, thorax, and abdomen—the head and thorax may be fused in some species to form a **cephalothorax**, which is covered by a plate called the carapace ([link]). The exoskeleton of many species is also infused with calcium carbonate, which makes it even stronger than in other arthropods. Crustaceans have an open circulatory system in which blood is pumped into the hemocoel by the dorsal heart. Most crustaceans typically have separate sexes, but some, like barnacles, may be hermaphroditic. Serial hermaphroditism, in which the gonad can switch from producing sperm to ova, is also found in some crustacean species. Larval stages are seen in the

early development of many crustaceans. Most crustaceans are carnivorous, but detritivores and filter feeders are also common.



The crayfish is an example of a crustacean. It has a carapace around the cephalothorax and the heart in the dorsal thorax area. (credit: Jane Whitney)

Subphylum Chelicerata includes animals such as spiders, scorpions, horseshoe crabs, and sea spiders. This subphylum is predominantly terrestrial, although some marine species also exist. An estimated 103,000[footnote] described species are included in subphylum Chelicerata. “Number of Living Species in Australia and the World,” A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

The body of chelicerates may be divided into two parts and a distinct “head” is not always discernible. The phylum derives its name from the first pair of appendages: the **chelicerae** ([link]**a**), which are specialized mouthparts. The chelicerae are mostly used for feeding, but in spiders, they are typically modified to inject venom into their prey ([link]**b**). As in other members of Arthropoda, chelicerates also utilize an open circulatory system, with a tube-like heart that pumps blood into the large hemocoel that bathes the internal organs. Aquatic chelicerates utilize gill respiration, whereas terrestrial species use either tracheae or book lungs for gaseous exchange.

(a) The chelicerae (first set of appendages) are well developed in the Chelicerata, which includes scorpions (a) and spiders (b). (credit a: modification of work by Kevin Walsh; credit b: modification of work by Marshal Hedin)

Note:
Concept in Action

[Click through](#) this lesson on arthropods to explore interactive habitat maps and more.

Section Summary

Flatworms are acoelomate, triploblastic animals. They lack circulatory and respiratory systems, and have a rudimentary excretory system. The digestive system is incomplete in most species. There are four traditional

classes of flatworms, the largely free-living turbellarians, the ectoparasitic monogeneans, and the endoparasitic trematodes and cestodes. Trematodes have complex life cycles involving a secondary mollusk host and a primary host in which sexual reproduction takes place. Cestodes, or tapeworms, infect the digestive systems of primary vertebrate hosts.

Nematodes are pseudocoelomate members of the clade Ecdysozoa. They have a complete digestive system and a pseudocoelomic body cavity. This phylum includes free-living as well as parasitic organisms. They include dioecious and hermaphroditic species. Nematodes have a poorly developed excretory system. Embryonic development is external and proceeds through larval stages separated by molts.

Arthropods represent the most successful phylum of animals on Earth, in terms of number of species as well as the number of individuals. They are characterized by a segmented body and jointed appendages. In the basic body plan, a pair of appendages is present per body segment. Within the phylum, classification is based on mouthparts, number of appendages, and modifications of appendages. Arthropods bear a chitinous exoskeleton. Gills, tracheae, and book lungs facilitate respiration. Embryonic development may include multiple larval stages.

Review Questions

Exercise:

Problem:

Which group of flatworms are primarily external parasites of fish?

a. monogeneans
b. trematodes
c. cestodes
d. turbellarians

Solution:

A

Exercise:

Problem: Crustaceans are _____.

a. ecdysozoans
b. nematodes
c. arachnids
d. parazoans

Solution:

A

Free Response

Exercise:

Problem:

Speculate as to what advantage(s) a complete digestive system has over an incomplete digestive system?

Solution:

In a complete digestive system, food material is not mixed with waste material, so the digestion and uptake of nutrients can be more efficient. In addition, the complete digestive system allows for an orderly progression of digestion of food matter and the specialization of different zones of the digestive tract.

Exercise:

Problem:

Describe a potential advantage and disadvantage of the cuticle of ecdysozoans.

Solution:

An advantage is that it is a tough covering that is protective against adverse environments, and predators and parasites. A disadvantage is that it must be shed and regrown for the animal to grow, which requires energy and makes the animal vulnerable during this process.

Glossary

Arthropoda
: a phylum of Ecdysozoa with jointed appendages and segmented bodies

cephalothorax
: a fused head and thorax

chelicerae
: a modified first pair of appendages in subphylum Chelicerata

chitin
: a tough nitrogen-containing polysaccharide found in the cuticles of arthropods and the cell walls of fungi

complete digestive system
: a digestive system that opens at one end, the mouth, and exits at the other end, the anus, and through which food normally moves in one direction

dioecious
: having separate male and female sexes

hemocoel
: the internal body cavity seen in arthropods

Nematoda
: a phylum of worms in Ecdysozoa commonly called roundworms containing both free-living and parasitic forms

spiracle
: a respiratory openings in insects that allow air into the tracheae

trachea
: in some arthropods, such as insects, a respiratory tube that conducts air from the spiracles to the tissues

Mollusks and Annelids

By the end of this section, you will be able to:

- Describe the unique anatomical features of mollusks
- Describe the features of an animal classified in phylum Annelida

The mollusks are a diverse group (85,000 described species) of mostly marine species. They have a variety of forms, ranging from large predatory squid and octopus, some of which show a high degree of intelligence, to small grazing forms with elaborately sculpted and colored shells. The annelids traditionally include the oligochaetes, which include the earthworms and leeches, the polychaetes, which are a marine group, and two other smaller classes.

The phyla Mollusca and Annelida belong to a clade called the **Lophotrochozoa**, which also includes the phylum Nemertea, or ribbon worms ([link]). They are distinct from the Ecdysozoa (nematodes and arthropods) based on evidence from analysis of their DNA, which has changed our views of the relationships among invertebrates.

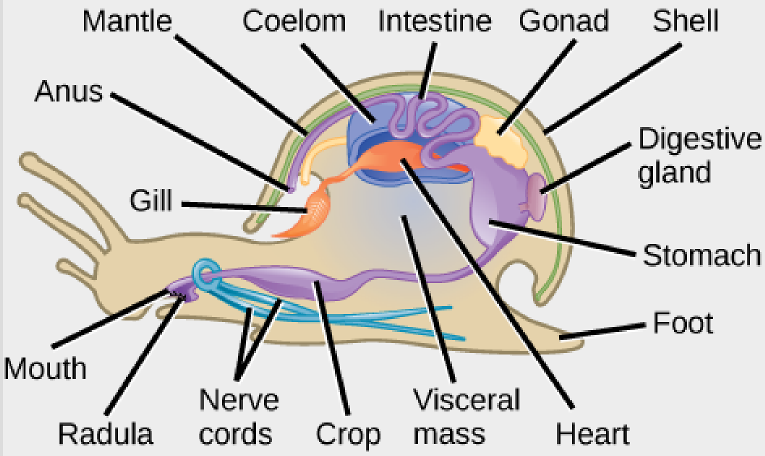
Phylum Mollusca

Mollusca is the predominant phylum in marine environments, where it is estimated that 23 percent of all known marine species belong to this phylum. It is the second most diverse phylum of animals with over 75,000 described species. The name "mollusca" signifies a soft body, as the earliest descriptions of mollusks came from observations of unshelled, soft-bodied cuttlefish (squid relatives). Although mollusk body forms vary, they share key characteristics, such as a ventral, muscular foot that is typically used for locomotion; the visceral mass, which contains most of the internal organs of the animal; and a dorsal mantle, which is a flap of tissue over the visceral mass that creates a space called the mantle cavity. The mantle may or may not secrete a shell of calcium carbonate. In addition, many mollusks have a scraping structure at the mouth, called a **radula** ([link]).

The muscular foot varies in shape and function, depending on the type of mollusk (described below in the section on mollusk diversity). It is a

retractable as well as extendable organ, used for locomotion and anchorage. Mollusks are eucoelomates, but the coelomic cavity is restricted to a cavity around the heart in adult animals. The mantle cavity, formed inside the **mantle**, develops independently of the coelomic cavity. It is a multi-purpose space, housing the gills, the anus, organs for sensing food particles in the water, and an outlet for gametes. Most mollusks have an open circulatory system with a heart that circulates the hemolymph in open spaces around the organs. The octopuses and squid are an exception to this and have a closed circulatory system with two hearts that move blood through the gills and a third, systemic heart that pumps blood through the rest of the body.

Note:
Art Connection



There are many species and variations of mollusks; the gastropod mollusk anatomy is shown here, which shares many characteristics common with other groups.

Which of the following statements about the anatomy of a mollusk is false?

a. Mollusks have a radula for scraping food.

b. Mollusks have ventral nerve cords.
c. The tissue beneath the shell is called the mantle.
d. The mantle cavity contains hemolymph.

Mollusk Diversity

This phylum is comprised of seven classes: Aplacophora, Monoplacophora, Polyplacophora, Bivalvia, Gastropoda, Cephalopoda, and Scaphopoda.

Class Aplacophora ("bearing no plates") includes worm-like animals living mostly on deep ocean bottoms. These animals lack a shell but have aragonite spicules on their skin. Members of class Monoplacophora ("bearing one plate") have a single, cap-like shell enclosing the body. The monoplacophorans were believed extinct and only known as fossils until the discovery of *Neopilina galatheae* in 1952. Today, scientists have identified nearly two dozen living species.

Animals in the class Polyplacophora ("bearing many plates") are commonly known as "chitons" and bear an armor-like, eight-plated shell ([link]). These animals have a broad, ventral foot that is adapted for attachment to rocks and a mantle that extends beyond the shell in the form of a girdle. They breathe with **ctenidia** (gills) present ventrally. These animals have a radula modified for scraping. A single pair of nephridia for excretion is present.

This chiton from the class Polyplacophora has the eight-plated shell indicative of its class. (credit: Jerry Kirkhart)

Class Bivalvia ("two shells") includes clams, oysters, mussels, scallops, and geoducks. They are found in marine and freshwater habitats. As the name suggests, bivalves are enclosed in a pair of shells (or valves) that are hinged at the dorsal side. The body is flattened on the sides. They feed by filtering particles from water and a radula is absent. They exchange gases using a pair of ctenidia, and excretion and osmoregulation are carried out by a pair of nephridia. In some species, the posterior edges of the mantle may fuse to form two siphons that inhale and exhale water. Some bivalves like oysters and mussels have the unique ability to secrete and deposit a calcareous **nacre** or "mother of pearl" around foreign particles that enter the mantle cavity. This property is commercially exploited to produce pearls.

Note:
Concept in Action

Watch animations of clams and mussels feeding to understand more about bivalves.

Gastropods (“stomach foot”) include well-known mollusks like snails, slugs, conchs, sea hares, and sea butterflies. Gastropods include shell-bearing species as well as species with a reduced shell. These animals are asymmetrical and usually present a coiled shell ([link]).

(a) (b)

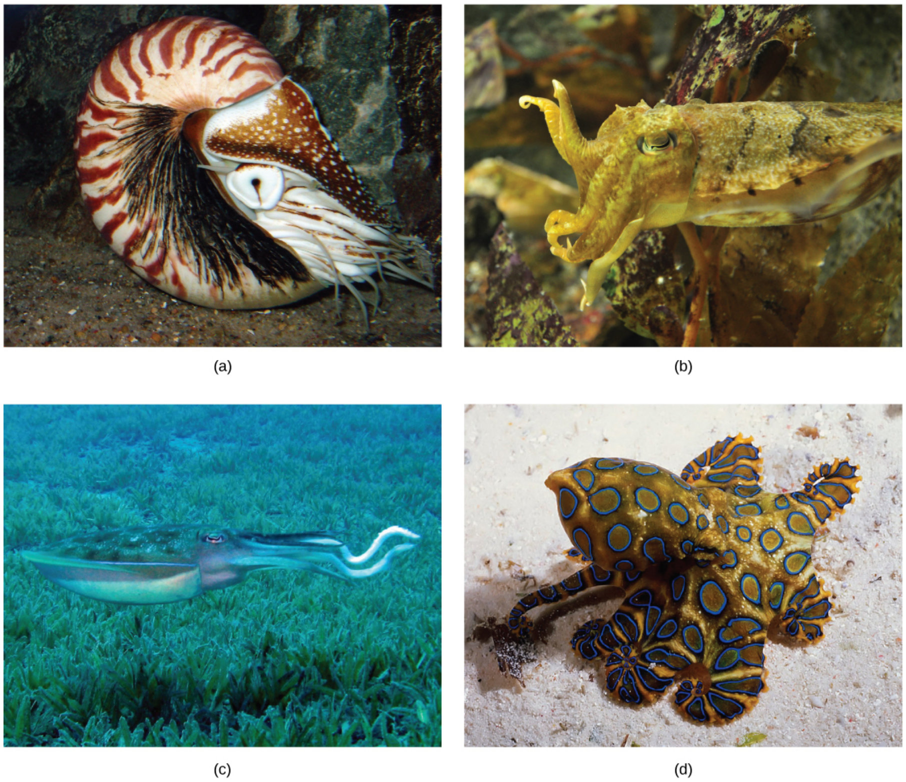
(a) Like many gastropods, this snail has a stomach foot and a coiled shell. (b) This slug, which is also a gastropod, lacks a shell. (credit a: modification of work by Murray Stevenson; credit b: modification of work by Rosendahl)

The visceral mass in the shelled species is characteristically twisted and the foot is modified for crawling. Most gastropods bear a head with tentacles

that support eyes. A complex radula is used to scrape food particles from the substrate. The mantle cavity encloses the ctenidia as well as a pair of nephridia.

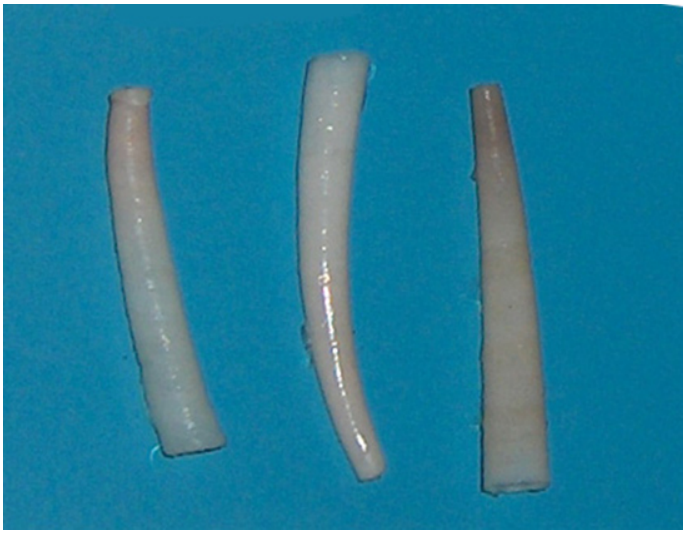
The class Cephalopoda (“head foot” animals) includes octopuses, squids, cuttlefish, and nautilus. Cephalopods include shelled and reduced-shell groups. They display vivid coloration, typically seen in squids and octopuses, which is used for camouflage. The ability of some octopuses to rapidly adjust their colors to mimic a background pattern or to startle a predator is one of the more awe-inspiring feats of these animals. All animals in this class are predators and have beak-like jaws. All cephalopods have a well-developed nervous system, complex eyes, and a closed circulatory system. The foot is lobed and developed into tentacles and a funnel, which is used for locomotion. Suckers are present on the tentacles in octopuses and squid. Ctenidia are enclosed in a large mantle cavity and are serviced by large blood vessels, each with its own heart.

Cephalopods ([link]) are able to move quickly via jet propulsion by contracting the mantle cavity to forcefully eject a stream of water. Cephalopods have separate sexes, and the females of some species care for the eggs for an extended period of time. Although the shell is much reduced and internal in squid and cuttlefish, and absent altogether in octopus, nautilus live inside a spiral, multi-chambered shell that is filled with gas or water to regulate buoyancy.



The (a) nautilus, (b) giant cuttlefish, (c) reef squid, and (d) blue-ring octopus are all members of the class Cephalopoda. (credit a: modification of work by J. Baecker; credit b: modification of work by Adrian Mohedano; credit c: modification of work by Silke Baron; credit d: modification of work by Angell Williams)

Members of the class Scaphopoda (“boat feet”) are known colloquially as “tusk shells” or “tooth shells.” Tooth shells are open at both ends and usually lie buried in sand with the front opening exposed to water and the reduced head end projecting from the back of the shell. Tooth shells have a radula and a foot modified into tentacles, each with a bulbous end that catches and manipulates prey ([link]).



Antalis vulgaris shows the classic Dentaliidae shape that gives these animals their common name of “tusk shell.” (credit: Georges Jansoone)

Annelida

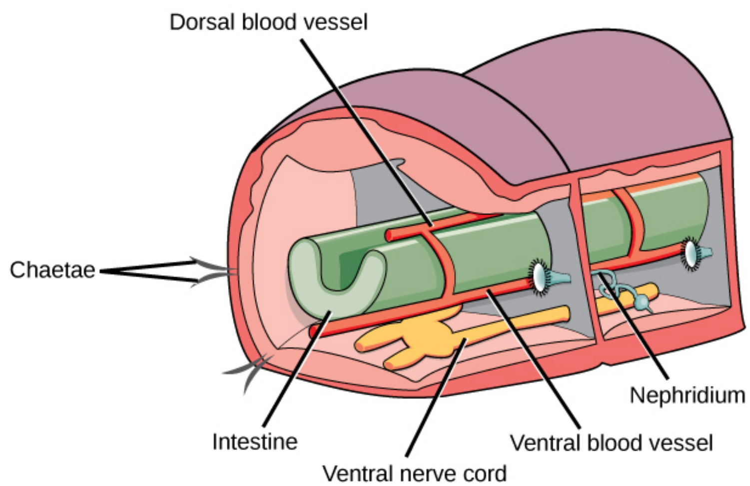
Phylum **Annelida** are segmented worms found in marine, terrestrial, and freshwater habitats, but the presence of water or humidity is a critical factor for their survival in terrestrial habitats. The name of the phylum is derived from the Latin word *annellus*, which means a small ring. Approximately 16,500 species have been described. The phylum includes earthworms, polychaete worms, and leeches. Like mollusks, annelids exhibit protostomic development.

Annelids are bilaterally symmetrical and have a worm-like appearance. Their particular segmented body plan results in repetition of internal and external features in each body segment. This type of body plan is called **metamerism**. The evolutionary benefit of such a body plan is thought to be the capacity it allows for the evolution of independent modifications in different segments that perform different functions. The overall body can then be divided into head, body, and tail.

Physiological Processes of Annelida

The skin of annelids is protected by a cuticle that is thinner than the cuticle of the ecdysozoans and does not need to be molted for growth. Chitinous hairlike extensions, anchored in the skin and projecting from the cuticle, called **chaetae**, are present in every segment in most groups. The chaetae are a defining character of annelids. Polychaete worms have paired, unjointed limbs called parapodia on each segment used for locomotion and breathing. Beneath the cuticle there are two layers of muscle, one running around its circumference (circular) and one running the length of the worm (longitudinal). Annelids have a true coelom in which organs are distributed and bathed in coelomic fluid. Annelids possess a well-developed complete digestive system with specialized organs: mouth, muscular pharynx, esophagus, and crop. A cross-sectional view of a body segment of an earthworm is shown in [link]; each segment is limited by a membrane that divides the body cavity into compartments.

Annelids have a closed circulatory system with muscular pumping “hearts” in the anterior segments, dorsal and ventral blood vessels that run the length of the body with connections in each segment, and capillaries that service individual tissues. Gas exchange occurs across the moist body surface. Excretion is carried out by pairs of primitive “kidneys” called metanephridia that consist of a convoluted tubule and an open, ciliated funnel present in every segment. Annelids have a well-developed nervous system with two ventral nerve cords and a nerve ring of fused ganglia present around the pharynx.



In this schematic showing the basic anatomy of annelids, the digestive system is indicated in green, the nervous system is indicated in yellow, and the circulatory system is indicated in red.

Annelids may be either monoecious with permanent gonads (as in earthworms and leeches) or dioecious with temporary or seasonal gonads (as in polychaetes).

Note:
Concept in Action

This video and animation provides a close-up look at annelid anatomy.

Annelid Diversity

Phylum Annelida includes the classes Polychaeta and Clitellata ([link]); the latter contains subclasses Oligochaeta, Hirudinoidea, and Branchiobdellida.

Earthworms are the most abundant members of the subclass Oligochaeta, distinguished by the presence of the **clitellum**, a ring structure in the skin that secretes mucus to bind mating individuals and forms a protective cocoon for the eggs. They also have a few, reduced chaetae (oligo- = “few”; -chaetae = “hairs”). The number and size of chaetae is greatly diminished in oligochaetes as compared to the polychaetes (poly- = “many”; -chaetae = “hairs”). The chaetae of polychaetes are also arranged within fleshy, flat, paired appendages on each segment called parapodia.

The subclass Hirudinoidea includes leeches. Significant differences between leeches and other annelids include the development of suckers at the anterior and posterior ends, and the absence of chaetae. Additionally, the segmentation of the body wall may not correspond to internal segmentation of the coelomic cavity. This adaptation may allow leeches to swell when ingesting blood from host vertebrates. The subclass Branchiobdellida includes about 150 species that show similarity to leeches as well as oligochaetes. All species are obligate symbionts, meaning that they can only survive associated with their host, mainly with freshwater crayfish. They feed on the algae that grows on the carapace of the crayfish.

(a) (b)

The (a) earthworm and (b) leech are both annelids.

(credit a: modification of work by "schizoform"/Flickr; credit b: modification of work by "Sarah G..."/Flickr)

Section Summary

The phylum Mollusca is a large, mainly marine group of invertebrates. Mollusks show a variety of morphologies. Many mollusks secrete a calcareous shell for protection, but in other species, the shell is reduced or absent. Mollusks are protostomes. The dorsal epidermis in mollusks is modified to form the mantle, which encloses the mantle cavity and visceral organs. This cavity is distinct from the coelomic cavity, which the adult animal retains, surrounding the heart. Respiration is facilitated by gills known as ctenidia. A chitinous scraper called the radula is present in most mollusks. Mollusks are mostly dioecious and are divided into seven classes.

The phylum Annelida includes worm-like, segmented animals. Segmentation is both external and internal, which is called metamerism. Annelids are protostomes. The presence of chitinous hairs called chaetae is characteristic of most members. These animals have well-developed nervous and digestive systems. Polychaete annelids have parapodia that participate in locomotion and respiration. Suckers are seen in the order Hirudinea. Breeding systems include separate sexes and hermaphroditism.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the anatomy of a mollusk is false?

a. Mollusks have a radula for scraping food.
b. Mollusks have ventral nerve cords.
c. The tissue beneath the shell is called the mantle.

d. The mantle cavity contains hemolymph.

Solution:

[link] D

Review Questions

Exercise:

Problem:A mantle and mantle cavity are present in ______.

a. class Oligochaeta
b. class Bivalvia
c. class Polychaeta
d. class Hirudinea

Solution:

C

Exercise:

Problem: Annelids have a ______.

a. pseudocoelom
b. a true coelom
c. no coelom
d. none of the above

Solution:

B

Free Response

Exercise:

Problem: Describe the morphology and anatomy of mollusks.

Solution:

Mollusks have a large muscular foot that may be modified in various ways, such as into tentacles, but it functions in locomotion. They have a mantle, a structure of tissue that covers and encloses the dorsal portion of the animal and secretes the shell when it is present. The mantle encloses the mantle cavity, which houses the gills (when present), excretory pores, anus, and gonadopores. The coelom of mollusks is restricted to the region around the systemic heart. The main body cavity is a hemocoel. Many mollusks have a radula near the mouth that is used for scraping food.

Glossary

Annelida
: a phylum of worm-like animals with metamerism

chaeta
: a chitinous projection from the cuticle found in annelids

clitellum
: a specialized band of fused segments in some annelids, which aids in reproduction

ctenidia
: specialized gills in mollusks

Lophotrochozoa
: a clade of invertebrate organisms that is a sister group to the Ecdysozoa

mantle
: a specialized epidermis that encloses all visceral organs and secretes shells in mollusks

metamerism
: having a series of body structures that are similar internally and externally, such as segments

Mollusca
: a phylum of protostomes with soft bodies and no segmentation

nacre
: a calcareous secretion produced by bivalve mollusks to line the inner side of shells as well as to coat foreign particulate matter

radula
: a tongue-like scraping organ with chitinous ornamentation found in most mollusks

Echinoderms and Chordates
By the end of this section, you will be able to:

- Describe the distinguishing characteristics of echinoderms
- Describe the distinguishing characteristics of chordates

Deuterostomes include the phyla Echinodermata and Chordata (which includes the vertebrates) and two smaller phyla. Deuterostomes share similar patterns of early development.

Echinoderms

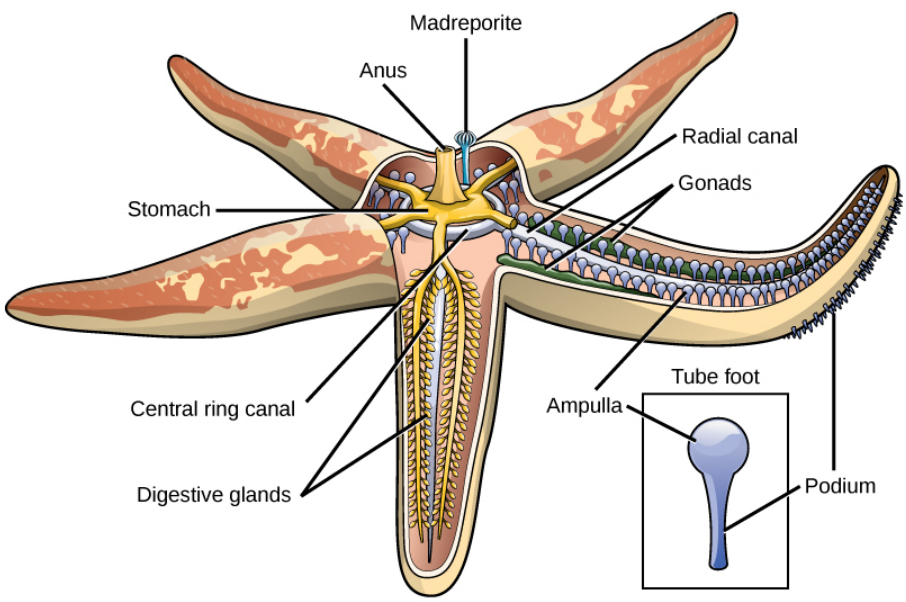
Echinodermata are named for their spiny skin (from the Greek "echinos" meaning "spiny" and "dermos" meaning "skin"). The phylum includes about 7,000[footnote] described living species, such as sea stars, sea cucumbers, sea urchins, sand dollars, and brittle stars. **Echinodermata** are exclusively marine.
"Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

Adult echinoderms exhibit pentaradial symmetry and have a calcareous endoskeleton made of ossicles ([link]), although the early larval stages of all echinoderms have bilateral symmetry. The endoskeleton is developed by epidermal cells, which may also possess pigment cells, giving vivid colors to these animals, as well as cells laden with toxins. These animals have a true coelom, a portion of which is modified into a unique circulatory system called a **water vascular system**. An interesting feature of these animals is their power to regenerate, even when over 75 percent of their body mass is lost.

Physiological Processes of Echinoderms

Echinoderms have a unique system for gas exchange, nutrient circulation, and locomotion called the water vascular system. The system consists of a central ring canal and radial canals extending along each arm. Water

circulates through these structures allowing for gas, nutrient, and waste exchange. A structure on top of the body, called the **madreporite**, regulates the amount of water in the water vascular system. “Tube feet,” which protrude through openings in the endoskeleton, may be expanded or contracted using the hydrostatic pressure in the system. The system allows for slow movement, but a great deal of power, as witnessed when the tube feet latch on to opposite halves of a bivalve mollusk, like a clam, and slowly, but surely pull the shells apart, exposing the flesh within.



This diagram shows the anatomy of a sea star.

The echinoderm nervous system has a nerve ring at the center and five radial nerves extending outward along the arms. There is no centralized nervous control. Echinoderms have separate sexes and release their gametes into the water where fertilization takes place. Echinoderms may also reproduce asexually through regeneration from body parts.

Echinoderm Diversity

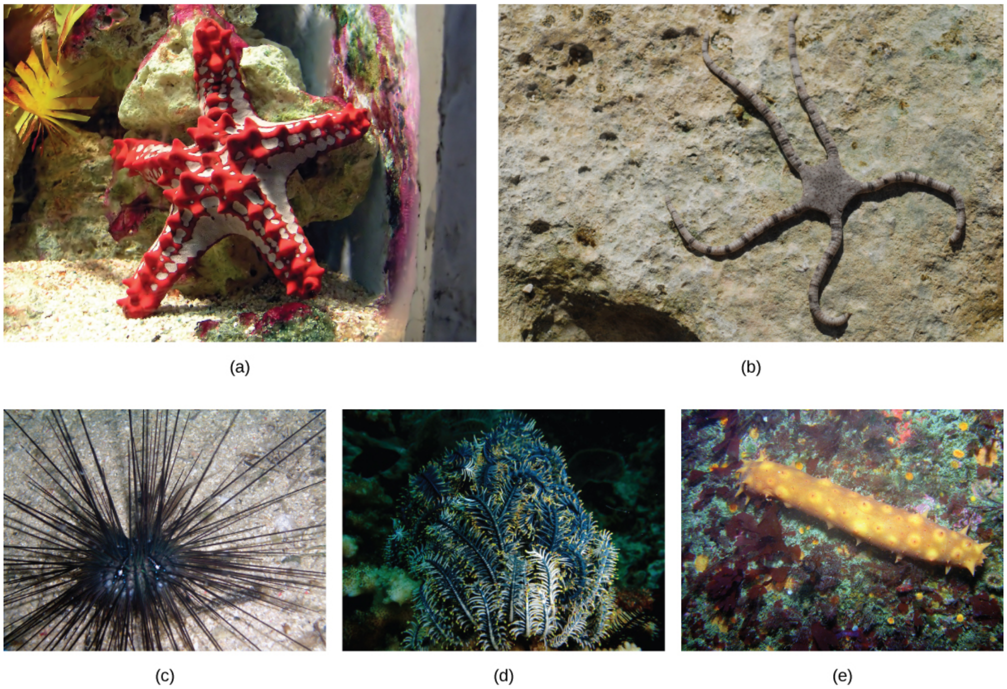
This phylum is divided into five classes: Asteroidea (sea stars), Ophiuroidea (brittle stars), Echinoidea (sea urchins and sand dollars), Crinoidea (sea lilies or feather stars), and Holothuroidea (sea cucumbers) ([link]).

Perhaps the best-known echinoderms are members of the class Asteroidea, or sea stars. They come in a large variety of shapes, colors, and sizes, with more than 1,800 species known. The characteristics of sea stars that set them apart from other echinoderm classes include thick arms that extend from a central disk where organs penetrate into the arms. Sea stars use their tube feet not only for gripping surfaces but also for grasping prey. Sea stars have two stomachs, one of which they can evert through their mouths to secrete digestive juices into or onto prey before ingestion. This process can essentially liquefy the prey and make digestion easier.

Note:
Concept in Action

View this video to explore a sea star’s body plan up close, watch one move across the sea floor, and see it devour a mussel.

Brittle stars have long, thin arms that do not contain any organs. Sea urchins and sand dollars do not have arms but are hemispherical or flattened with five rows of tube feet, which help them in slow movement. Sea lilies and feather stars are stalked suspension feeders. Sea cucumbers are soft-bodied and elongate with five rows of tube feet and a series of tube feet around the mouth that are modified into tentacles used in feeding.



Different members of Echinodermata include the (a) sea star in class Asteroidea, (b) the brittle star in class Ophiuroidea, (c) the sea urchins of class Echinoidea, (d) the sea lilies belonging to class Crinoidea, and (e) sea cucumbers representing class Holothuroidea. (credit a: modification of work by Adrian Pingstone; credit b: modification of work by Joshua Ganderson; credit c: modification of work by Samuel Chow; credit d: modification of work by Sarah Depper; credit e: modification of work by Ed Bierman)

Chordates

The majority of species in the phylum Chordata are found in the subphylum Vertebrata, which include many species with which we are familiar. The vertebrates contain more than 60,000 described species, divided into major groupings of the lampreys, fishes, amphibians, reptiles, birds, and mammals.

Animals in the phylum **Chordata** share four key features that appear at some stage of their development: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail ([link]). In certain groups, some of these traits are present only during embryonic development.

The chordates are named for the **notochord**, which is a flexible, rod-shaped structure that is found in the embryonic stage of all chordates and in the adult stage of some chordate species. It is located between the digestive tube and the nerve cord, and provides skeletal support through the length of the body. In some chordates, the notochord acts as the primary axial support of the body throughout the animal's lifetime. In vertebrates, the notochord is present during embryonic development, at which time it induces the development of the neural tube and serves as a support for the developing embryonic body. The notochord, however, is not found in the postnatal stage of vertebrates; at this point, it has been replaced by the **vertebral column** (the spine).

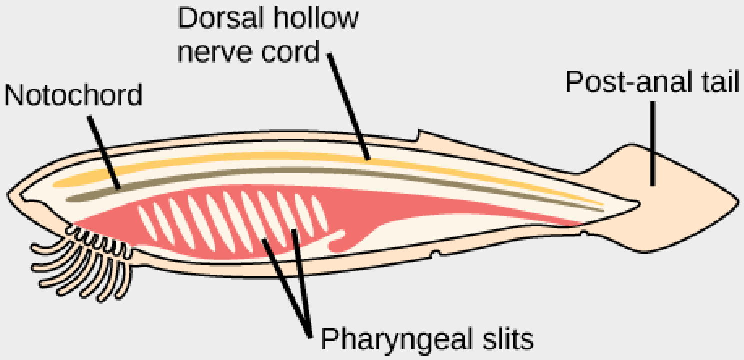
The **dorsal hollow nerve cord** is derived from ectoderm that sinks below the surface of the skin and rolls into a hollow tube during development. In chordates, it is located dorsally to the notochord. In contrast, other animal phyla possess solid nerve cords that are located either ventrally or laterally. The nerve cord found in most chordate embryos develops into the brain and spinal cord, which compose the central nervous system.

Pharyngeal slits are openings in the pharynx, the region just posterior to the mouth, that extend to the outside environment. In organisms that live in aquatic environments, pharyngeal slits allow for the exit of water that enters the mouth during feeding. Some invertebrate chordates use the pharyngeal slits to filter food from the water that enters the mouth. In fishes, the pharyngeal slits are modified into gill supports, and in jawed fishes, jaw supports. In tetrapods, the slits are further modified into components of the ear and tonsils, since there is no longer any need for gill supports in these air-breathing animals. ***Tetrapod*** means "four-footed," and this group includes amphibians, reptiles, birds, and mammals. (Birds are considered tetrapods because they evolved from tetrapod ancestors.)

The **post-anal tail** is a posterior elongation of the body extending beyond the anus. The tail contains skeletal elements and muscles, which provide a

source of locomotion in aquatic species, such as fishes. In some terrestrial vertebrates, the tail may also function in balance, locomotion, courting, and signaling when danger is near. In many species, the tail is absent or reduced; for example, in apes, including humans, it is present in the embryo, but reduced in size and nonfunctional in adults.

Note:
Art Connection



In chordates, four common features appear at some point in development: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail. The anatomy of a cephalochordate shown here illustrates all of these features.

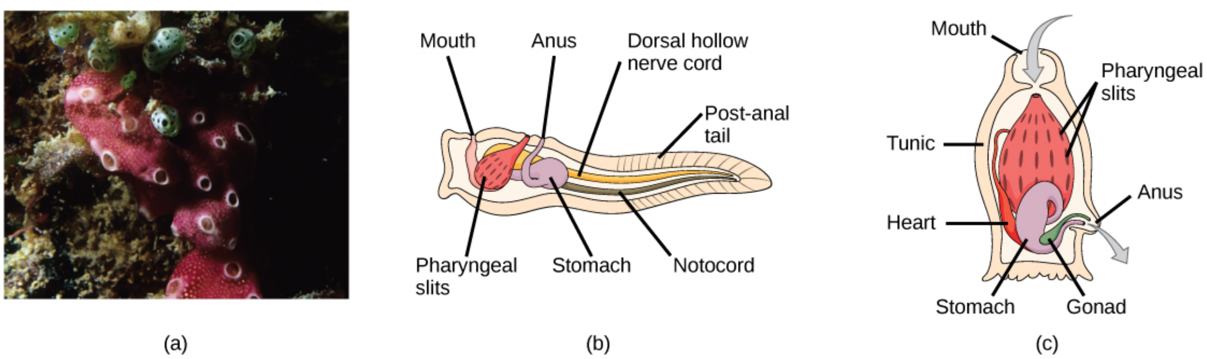
Which of the following statements about common features of chordates is true?

a. The dorsal hollow nerve cord is part of the chordate central nervous system.
b. In vertebrate fishes, the pharyngeal slits become the gills.
c. Humans are not chordates because humans do not have a tail.
d. Vertebrates do not have a notochord at any point in their development; instead, they have a vertebral column.

Invertebrate Chordates

In addition to the vertebrates, the phylum Chordata contains two clades of invertebrates: **Urochordata** (tunicates) and **Cephalochordata** (lancelets). Members of these groups possess the four distinctive features of chordates at some point during their development.

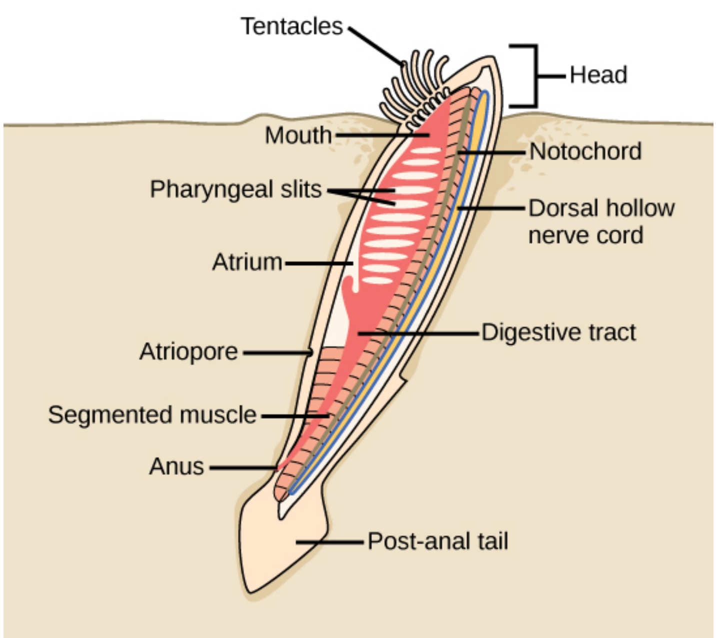
The **tunicates** ([link]) are also called sea squirts. The name tunicate derives from the cellulose-like carbohydrate material, called the tunic, which covers the outer body. Although tunicates are classified as chordates, the adult forms are much modified in body plan and do not have a notochord, a dorsal hollow nerve cord, or a post-anal tail, although they do have pharyngeal slits. The larval form possesses all four structures. Most tunicates are hermaphrodites. Tunicate larvae hatch from eggs inside the adult tunicate's body. After hatching, a tunicate larva swims for a few days until it finds a suitable surface on which it can attach, usually in a dark or shaded location. It then attaches by the head to the substrate and undergoes metamorphosis into the adult form, at which point the notochord, nerve cord, and tail disappear.



(a) This photograph shows a colony of the tunicate *Botrylloides violaceus*. In the (b) larval stage, the tunicate can swim freely until it attaches to a substrate to become (c) an adult. (credit a: modification of work by Dr. Dwayne Meadows, NOAA/NMFS/OPR)

Most tunicates live a sessile existence in shallow ocean waters and are suspension feeders. The primary foods of tunicates are plankton and detritus. Seawater enters the tunicate’s body through its incurrent siphon. Suspended material is filtered out of this water by a mucus net (pharyngeal slits) and is passed into the intestine through the action of cilia. The anus empties into the excurrent siphon, which expels wastes and water.

Lancelets possess a notochord, dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail in the adult stage ([link]). The notochord extends into the head, which gives the subphylum its name (Cephalochordata). Extinct fossils of this subphylum date to the middle of the Cambrian period (540–488 mya).The living forms, the lancelets, are named for their blade-like shape. Lancelets are only a few centimeters long and are usually found buried in sand at the bottom of warm temperate and tropical seas. Like tunicates, they are suspension feeders.



Adult lancelets retain the four key features of chordates: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail.

Section Summary

Echinoderms are deuterostome marine organisms. This phylum of animals bear a calcareous endoskeleton composed of ossicles covered by a spiny skin. Echinoderms possess a water-based circulatory system. The madreporite is the point of entry and exit for water for the water vascular system.

The characteristic features of Chordata are a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail. Chordata contains two clades of invertebrates: Urochordata (tunicates) and Cephalochordata (lancelets), together with the vertebrates. Most tunicates live on the ocean floor and are suspension feeders. Lancelets are suspension feeders that feed on phytoplankton and other microorganisms.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about common features of chordates is true?

a. The dorsal hollow nerve cord is part of the chordate central nervous system.
b. In vertebrate fishes, the pharyngeal slits become the gills.
c. Humans are not chordates because humans do not have a tail.
d. Vertebrates do not have a notochord at any point in their development; instead, they have a vertebral column.

Solution:

[link] A

Review Questions

Exercise:

Problem: Echinoderms in their larval state have ______.

a. triangular symmetry
b. radial symmetry
c. hexagonal symmetry
d. bilateral symmetry

Solution:

D

Exercise:

Problem: The circulatory fluid in echinoderms is ______.

a. blood
b. mesohyl
c. water
d. saline

Solution:

C

Exercise:

Problem:

Which of the following is *not* a member of the phylum Chordata?

a. Cephalochordata
b. Echinodermata
c. Urochordata

d. Vertebrata

Solution:

B

Free Response

Exercise:

Problem:

Sessile adult tunicates lose the notochord; what does this suggest about one function of this structure?

Solution:

It suggests that the notochord is important for support during locomotion of an organism.

Exercise:

Problem:

During embryonic development, what features do we share with tunicates or lancelets?

Solution:

During embryonic development, we also have a notochord, a dorsal hollow nerve tube, pharyngeal slits, and a post-anal tail.

Glossary

Cephalochordata
: a chordate clade whose members possess a notochord, dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail in the adult stage

Chordata
: a phylum of animals distinguished by their possession of a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail at some point during their development

dorsal hollow nerve cord
: a hollow, tubular structure derived from ectoderm, which is located dorsal to the notochord in chordates

Echinodermata
: a phylum of deuterostomes with spiny skin; exclusively marine organisms

lancelet
: a member of Cephalochordata; named for its blade-like shape

madreporite
: a pore for regulating entry and exit of water into the water vascular system

notochord
: a flexible, rod-shaped structure that is found in the embryonic stage of all chordates and in the adult stage of some chordates

pharyngeal slit
: an opening in the pharynx

post-anal tail
: a muscular, posterior elongation of the body extending beyond the anus in chordates

tetrapod
: a four-footed animal; includes amphibians, reptiles, birds, and mammals

tunicate
: a sessile chordate that is a member of Urochordata

Urochordata
the clade composed of the tunicates

vertebral column
a series of separate bones that surround the spinal cord in vertebrates

water vascular system
a system in echinoderms in which water is the circulatory fluid

Vertebrates

By the end of this section, you will be able to:

- Describe the difference between jawless and jawed fishes
- Explain the main characteristics of amphibians, reptiles, and birds
- Describe the derived characteristics in birds that facilitate flight
- Name and describe the distinguishing features of the three main groups of mammals
- Describe the derived features that distinguish primates from other animals

Vertebrates are among the most recognizable organisms of the animal kingdom ([link]). More than 62,000 vertebrate species have been identified. The vertebrate species now living represent only a small portion of the vertebrates that have existed. The best-known extinct vertebrates are the dinosaurs, a unique group of reptiles, reaching sizes not seen before or since in terrestrial animals. They were the dominant terrestrial animals for 150 million years, until they died out near the end of the Cretaceous period in a mass extinction. A great deal is known about the anatomy of the dinosaurs, given the preservation of their skeletal elements in the fossil record.

Examples of critically endangered vertebrate species include (a) the Siberian tiger *(Panthera tigris altaica*), (b) the Panamanian golden frog (*Atelopus zeteki*), and (c) the Philippine eagle (*Pithecophaga jefferyi*). (credit a: modification of work by Dave Pape; credit b: modification of work by Brian Gratwicke; credit c: modification of work by "cuatrok77"/Flickr)

Fishes

Modern fishes include an estimated 31,000 species. Fishes were the earliest vertebrates, and jawless fishes were the earliest of these. Jawless fishes—the present day hagfishes and lampreys—have a distinct cranium and complex sense organs including eyes, distinguishing them from the invertebrate chordates. The jawed fishes evolved later and are extraordinarily diverse today. Fishes are active feeders, rather than sessile, suspension feeders.

Jawless Fishes

Jawless fishes are **craniates** (which includes all the chordate groups except the tunicates and lancelets) that represent an ancient vertebrate lineage that arose over one half-billion years ago. Some of the earliest jawless fishes were the **ostracoderms** (which translates as "shell-skin"). Ostracoderms, now extinct, were vertebrate fishes encased in bony armor, unlike present-day jawless fishes, which lack bone in their scales.

The clade **Myxini** includes 67 species of hagfishes. **Hagfishes** are eel-like scavengers that live on the ocean floor and feed on dead invertebrates, other fishes, and marine mammals ([link]**a**). Hagfishes are entirely marine and are found in oceans around the world except for the polar regions. A unique feature of these animals is the slime glands beneath the skin that are able to release an extraordinary amount of mucus through surface pores. This mucus may allow the hagfish to escape from the grip of predators. Hagfish are known to enter the bodies of dead or dying organisms to devour them from the inside.

(a) (b)

(a) Pacific hagfishes are scavengers that live on the ocean floor. (b) These parasitic sea lampreys attach to their lake trout host by suction and use their rough tongues to rasp away flesh in order to feed on the trout’s blood. (credit a: modification of work by Linda Snook, NOAA/CBNMS; credit b: modification of work by USGS)

The skeleton of a hagfish is composed of cartilage, which includes a cartilaginous notochord, which runs the length of the body, and a skull. This notochord provides support to the fish’s body. Although they are craniates, hagfishes are not vertebrates, since they do not replace the notochord with a vertebral column during development, as do the vertebrates.

The clade **Petromyzontidae** includes approximately 40 species of lampreys. **Lampreys** are similar to hagfishes in size and shape; however, lampreys have a brain case and incomplete vertebrae. Lampreys lack paired appendages and bone, as do the hagfishes. As adults, lampreys are characterized by a toothed, funnel-like sucking mouth. Some species are parasitic as adults, attaching to and feeding on the body fluids of fish ([link]**b**). Most species are free-living.

Lampreys live primarily in coastal and fresh waters and have a worldwide temperate region distribution. All species spawn in fresh waters. Eggs are fertilized externally, and the larvae are distinctly different from the adult form, spending 3 to 15 years as suspension feeders. Once they attain sexual

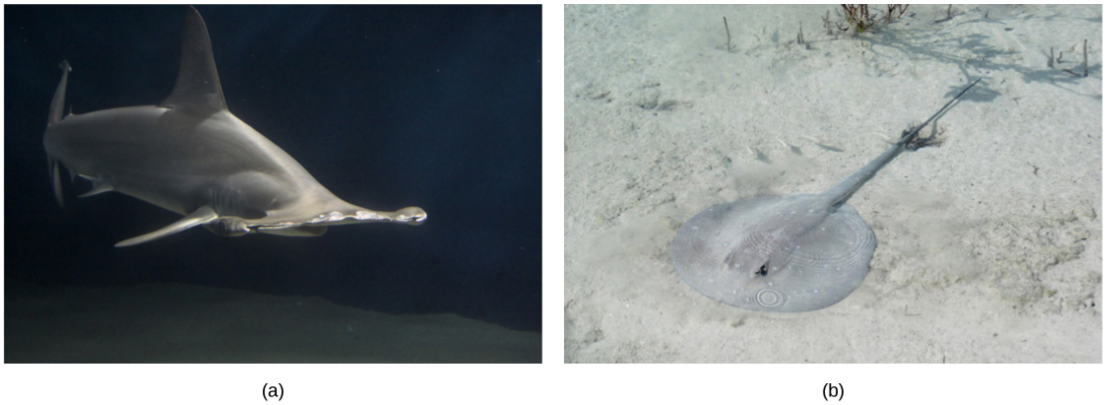
maturity, the adults reproduce and die within days. Lampreys have a notochord as adults.

Jawed Fishes

Gnathostomes or “jaw-mouths” are vertebrates that have jaws and include both cartilaginous and bony fishes. One of the most significant developments in early vertebrate evolution was the origin of the jaw, which is a hinged structure attached to the cranium that allows an animal to grasp and tear its food. The evolution of jaws allowed early gnathostomes to exploit food resources that were unavailable to jawless fishes.

The clade **Chondrichthyes**, the cartilaginous fishes, is diverse, consisting of sharks ([link]**a**), rays, and skates, together with sawfishes and a few dozen species of fishes called *chimaeras*, or ghost sharks. Chondrichthyes have paired fins and a skeleton made of cartilage. This clade arose approximately 370 million years ago in the middle Devonian. They are thought to have descended from an extinct group that had a skeleton made of bone; thus, the cartilaginous skeleton of Chondrichthyes is a later development. Parts of the shark skeleton are strengthened by granules of calcium carbonate, but this is not the same as bone.

Most cartilaginous fishes live in marine habitats, with a few species living in fresh water for some or all of their lives. Most sharks are carnivores that feed on live prey, either swallowing it whole or using their jaws and teeth to tear it into smaller pieces. Shark teeth likely evolved from the jagged scales that cover their skin. Some species of sharks and rays are suspension feeders that feed on plankton.



(a) This hammerhead shark is an example of a predatory cartilaginous fish. (b) This stingray blends into the sandy bottom of the ocean floor when it is feeding or awaiting prey. (credit a: modification of work by Masashi Sugawara; credit b: modification of work by "Sailn1"/Flickr)

Sharks have well-developed sense organs that aid them in locating prey, including a keen sense of smell and electroreception, the latter being perhaps the most sensitive of any animal. Organs called **ampullae of Lorenzini** allow sharks to detect the electromagnetic fields that are produced by all living things, including their prey. Electroreception has only been observed in aquatic or amphibious animals. Sharks, together with most fishes, also have a sense organ called the **lateral line**, which is used to detect movement and vibration in the surrounding water, and a sense that is often considered homologous to “hearing” in terrestrial vertebrates. The lateral line is visible as a darker stripe that runs along the length of the fish’s body.

Sharks reproduce sexually and eggs are fertilized internally. Most species are ovoviviparous, that is, the fertilized egg is retained in the oviduct of the mother’s body, and the embryo is nourished by the egg yolk. The eggs hatch in the uterus and young are born alive and fully functional. Some species of sharks are oviparous: They lay eggs that hatch outside of the mother’s body. Embryos are protected by a shark egg case or “mermaid’s purse” that has the consistency of leather. The shark egg case has tentacles

that snag in seaweed and give the newborn shark cover. A few species of sharks are viviparous, that is, the young develop within the mother's body, and she gives live birth.

Rays and skates include more than 500 species and are closely related to sharks. They can be distinguished from sharks by their flattened bodies, pectoral fins that are enlarged and fused to the head, and gill slits on their ventral surface ([link]**b**). Like sharks, rays and skates have a cartilaginous skeleton. Most species are marine and live on the sea floor, with nearly a worldwide distribution.

Bony Fishes

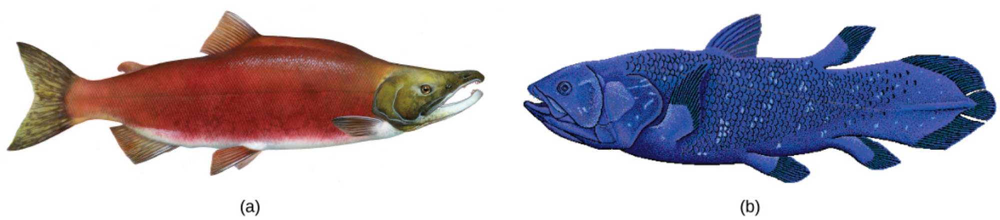
Members of the clade **Osteichthyes**, or bony fishes, are characterized by a bony skeleton. The vast majority of present-day fishes belong to this group, which consists of approximately 30,000 species, making it the largest class of vertebrates in existence today.

Nearly all bony fishes have an ossified skeleton with specialized bone cells (osteocytes) that produce and maintain a calcium phosphate matrix. This characteristic has only reverted in a few groups of Osteichthyes, such as sturgeons and paddlefish, which have primarily cartilaginous skeletons. The skin of bony fishes is often covered in overlapping scales, and glands in the skin secrete mucus that reduces drag when swimming and aids the fish in osmoregulation. Like sharks, bony fishes have a lateral line system that detects vibrations in water. Unlike sharks, some bony fish depend on their eyesight to locate prey. Bony fish are also unusual in possessing taste cells in the head and trunk region of the body that allow them to detect extremely small concentrations of molecules in the water.

All bony fishes, like the cartilaginous fishes, use gills to breathe. Water is drawn over gills that are located in chambers covered and ventilated by a protective, muscular flap called the operculum. Unlike sharks, bony fishes have a **swim bladder**, a gas-filled organ that helps to control the buoyancy of the fish. Bony fishes are further divided into two clades with living

members: **Actinopterygii** (ray-finned fishes) and **Sarcopterygii** (lobe-finned fishes).

The ray-finned fishes include many familiar fishes—tuna, bass, trout, and salmon ([link]**a**), among others. Ray-finned fishes are named for the form of their fins—webs of skin supported by bony spines called rays. In contrast, the fins of lobe-finned fishes are fleshy and supported by bone ([link]**b**). Living members of lobe-finned fishes include the less familiar lungfishes and coelacanth.



The (a) sockeye salmon and (b) coelacanth are both bony fishes of the Osteichthyes clade. The coelacanth, sometimes called a lobe-finned fish, was thought to have gone extinct in the Late Cretaceous period 100 million years ago until one was discovered in 1938 between Africa and Madagascar. (credit a: modification of work by Timothy Knepp, USFWS; credit b: modification of work by Robbie Cada)

Amphibians

Amphibians are vertebrate tetrapods. **Amphibia** includes frogs, salamanders, and caecilians. The term amphibian means “dual life,” which is a reference to the metamorphosis that many frogs undergo from a tadpole to an adult and the mixture of aquatic and terrestrial environments in their life cycle. Amphibians evolved in the Devonian period and were the earliest terrestrial tetrapods.

As tetrapods, most amphibians are characterized by four well-developed limbs, although some species of salamanders and all caecilians possess only vestigial limbs. An important characteristic of extant amphibians is a moist, permeable skin, achieved by mucus glands. The moist skin allows oxygen and carbon dioxide exchange with the environment, a process called **cutaneous respiration**. All living adult amphibian species are carnivorous, and some terrestrial amphibians have a sticky tongue that is used to capture prey.

Amphibian Diversity

Amphibia comprise an estimated 6,500 extant species that inhabit tropical and temperate regions around the world. Amphibians can be divided into three clades: **Urodela** (“tailed-ones”), the salamanders and newts; **Anura** (“tail-less ones”), the frogs and toads; and **Apoda** (“legless ones”), the caecilians.

Living **salamanders** ([link]**a**) include approximately 500 species, some of which are aquatic, others terrestrial, and some that live on land only as adults. Adult salamanders usually have a generalized tetrapod body plan with four limbs and a tail. Some salamanders are lungless, and respiration occurs through the skin or external gills. Some terrestrial salamanders have primitive lungs; a few species have both gills and lungs.

(a) (b)

(a) Most salamanders have legs and a tail, but respiration

varies among species. (b) The Australian green tree frog is a nocturnal predator that lives in the canopies of trees near a water source. (credit a: modification of work by Valentina Storti; credit b: modification of work by Evan Pickett)

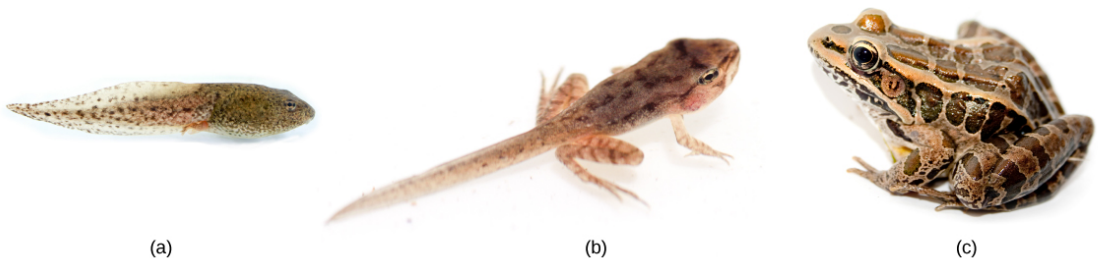
Note:
Concept in Action

Watch this video about an unusually large salamander species.

Frogs ([link]**b**) are the most diverse group of amphibians, with approximately 5,000 species that live on all continents except Antarctica. Frogs have a body plan that is more specialized than the salamander body plan for movement on land. Adult frogs use their hind limbs to jump many times their body length on land. Frogs have a number of modifications that allow them to avoid predators, including skin that acts as camouflage and defensive chemicals that are poisonous to predators secreted from glands in the skin.

Frog eggs are fertilized externally, as they are laid in moist environments. Frogs demonstrate a range of parental behaviors, with some species exhibiting little care, to species that carry eggs and tadpoles on their hind legs or backs. The life cycle consists of two stages: the larval stage followed by metamorphosis to an adult stage. The larval stage of a frog, the **tadpole**, is often a filter-feeding herbivore. Tadpoles usually have gills, a

lateral line system, long-finned tails, but no limbs. At the end of the tadpole stage, frogs undergo a gradual metamorphosis into the adult form. During this stage, the gills and lateral line system disappear, and four limbs develop. The jaws become larger and are suited for carnivorous feeding, and the digestive system transforms into the typical short gut of a predator. An eardrum and air-breathing lungs also develop. These changes during metamorphosis allow the larvae to move onto land in the adult stage ([link]).



A frog begins as a (a) tadpole and undergoes metamorphosis to become (b) a juvenile and finally (c) an adult. (credit: modification of work by Brian Gratwicke)

Caecilians comprise an estimated 185 species. They lack external limbs and resemble giant earthworms. They inhabit soil and are found primarily in the tropics of South America, Africa, and southern Asia where they are adapted for a soil-burrowing lifestyle and are nearly blind. Unlike most of the other amphibians that breed in or near water, reproduction in a drier soil habitat means that caecilians must utilize internal fertilization, and most species give birth to live young ([link]).

Caecilians lack external limbs and are well adapted for a soil-burrowing lifestyle. (credit: modification of work by "cliff1066"/Flickr)

Reptiles and Birds

The **amniotes**—reptiles, birds, and mammals—are distinguished from amphibians by their terrestrially adapted (shelled) egg and an embryo protected by amniotic membranes. The evolution of amniotic membranes meant that the embryos of amniotes could develop within an aquatic environment inside the egg. This led to less dependence on a water environment for development and allowed the amniotes to invade drier areas. This was a significant evolutionary change that distinguished them from amphibians, which were restricted to moist environments due to their shell-less eggs. Although the shells of various amniotic species vary significantly, they all allow retention of water. The membranes of the amniotic egg also allowed gas exchange and sequestering of wastes within the enclosure of an eggshell. The shells of bird eggs are composed of calcium carbonate and are hard and brittle, but possess pores for gas and water exchange. The shells of reptile eggs are more leathery and pliable. Most mammals do not lay eggs; however, even with internal gestation, amniotic membranes are still present.

In the past, the most common division of amniotes has been into classes Mammalia, Reptilia, and Aves. Birds are descended, however, from

dinosaurs, so this classical scheme results in groups that are not true clades. We will discuss birds as a group distinct from reptiles with the understanding that this does not reflect evolutionary history.

Reptiles

Reptiles are tetrapods. Limbless reptiles—snakes—may have vestigial limbs and, like caecilians, are classified as tetrapods because they are descended from four-limbed ancestors. Reptiles lay shelled eggs on land. Even aquatic reptiles, like sea turtles, return to the land to lay eggs. They usually reproduce sexually with internal fertilization. Some species display ovoviviparity, with the eggs remaining in the mother's body until they are ready to hatch. Other species are viviparous, with the offspring born alive.

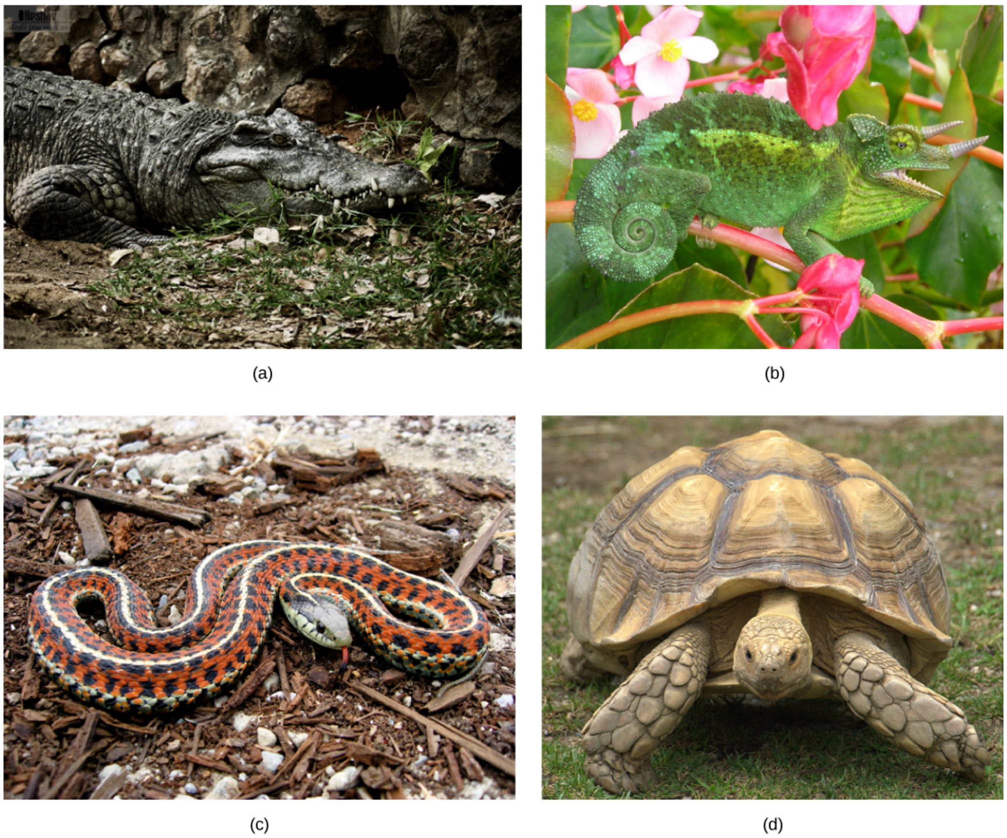
One of the key adaptations that permitted reptiles to live on land was the development of their scaly skin, containing the protein keratin and waxy lipids, which prevented water loss from the skin. This occlusive skin means that reptiles cannot use their skin for respiration, like amphibians, and thus all must breathe with lungs. In addition, reptiles conserve valuable body water by excreting nitrogen in the form of uric acid paste. These characteristics, along with the shelled, amniotic egg, were the major reasons why reptiles became so successful in colonizing a variety of terrestrial habitats far from water.

Reptiles are ectotherms, that is, animals whose main source of body heat comes from the environment. Behavioral maneuvers, like basking to heat themselves, or seeking shade or burrows to cool off, help them regulate their body temperature,

Class Reptilia includes diverse species classified into four living clades. These are the Crocodilia, Sphenodontia, Squamata, and Testudines.

The **Crocodilia** ("small lizard") arose approximately 84 million years ago, and living species include alligators, crocodiles, and caimans. Crocodilians ([link]**a**) live throughout the tropics of Africa, South America, the southeastern United States, Asia, and Australia. They are found in

freshwater habitats, such as rivers and lakes, and spend most of their time in water. Some species are able to move on land due to their semi-erect posture.



(a) Crocodilians, such as this Siamese crocodile, provide parental care for their offspring. (b) This Jackson’s chameleon blends in with its surroundings. (c) The garter snake belongs to the genus *Thamnophis*, the most widely distributed reptile genus in North America. (d) The African spurred tortoise lives at the southern edge of the Sahara Desert. It is the third largest tortoise in the world. (credit a: modification of work by Keshav Mukund Kandhadai; credit c: modification of work by Steve Jurvetson; credit d: modification of work by Jim Bowen)

The **Sphenodontia** (“wedge tooth”) arose in the Mesozoic Era and includes only one living genus, *Tuatara*, with two species that are found in New Zealand. There are many fossil species extending back to the Triassic period (250–200 million years ago). Although the tuataras resemble lizards, they are anatomically distinct and share characteristics that are found in birds and turtles.

Squamata (“scaly”) arose in the late Permian; living species include lizards and snakes, which are the largest extant clade of reptiles ([link]**b**). Lizards differ from snakes by having four limbs, eyelids, and external ears, which are lacking in snakes. Lizard species range in size from chameleons and geckos that are a few centimeters in length to the Komodo dragon, which is about 3 meters in length.

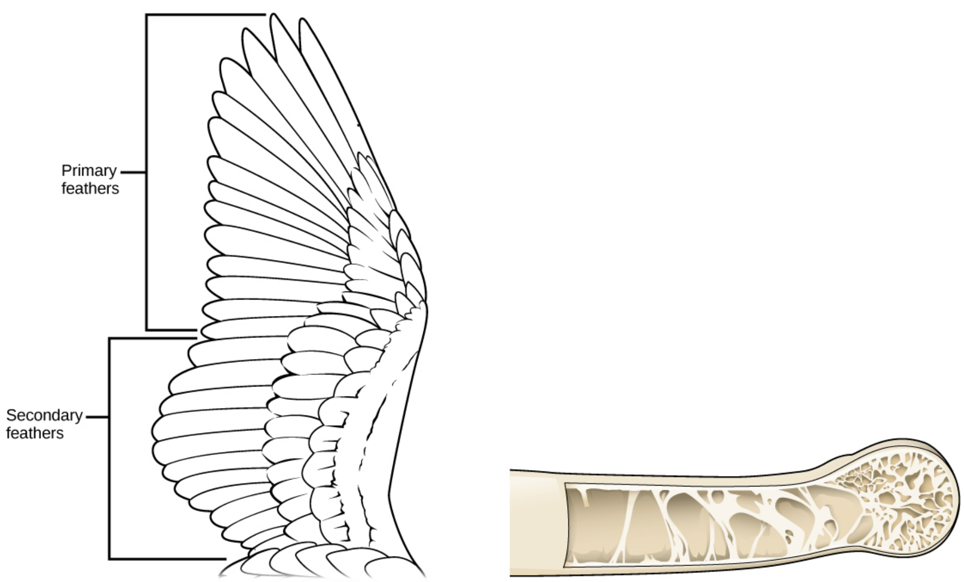
Snakes are thought to have descended from either burrowing lizards or aquatic lizards over 100 million years ago ([link]**c**). Snakes comprise about 3,000 species and are found on every continent except Antarctica. They range in size from 10 centimeter-long thread snakes to 7.5 meter-long pythons and anacondas. All snakes are carnivorous and eat small animals, birds, eggs, fish, and insects.

Turtles are members of the clade **Testudines** (“having a shell”) ([link]**d**). Turtles are characterized by a bony or cartilaginous shell, made up of the carapace on the back and the plastron on the ventral surface, which develops from the ribs. Turtles arose approximately 200 million years ago, predating crocodiles, lizards, and snakes. Turtles lay eggs on land, although many species live in or near water. Turtles range in size from the speckled padloper tortoise at 8 centimeters (3.1 inches) to the leatherback sea turtle at 200 centimeters (over 6 feet). The term “turtle” is sometimes used to describe only those species of Testudines that live in the sea, with the terms “tortoise” and “terrapin” used to refer to species that live on land and in fresh water, respectively.

Birds

Data now suggest that birds belong within the reptile clade, but they display a number of unique adaptations that set them apart. Unlike the reptiles, birds are endothermic, meaning they generate their own body heat through metabolic processes. The most distinctive characteristic of birds is their feathers, which are modified reptilian scales. Birds have several different types of feathers that are specialized for specific functions, like contour feathers that streamline the bird's exterior and loosely structured **down feathers** that insulate ([link]**a**).

Feathers not only permitted the earliest birds to glide, and ultimately engage in flapping flight, but they insulated the bird's body, assisting the maintenance of endothermy, even in cooler temperatures. Powering a flying animal requires economizing on the amount of weight carried. As body weight increases, the muscle output and energetic cost required for flying increase. Birds have made several modifications to reduce body weight, including hollow or **pneumatic bones** ([link]**b**) with air spaces that may be connected to air sacs and cross-linked struts within their bones to provide structural reinforcement. Parts of the vertebral skeleton and braincase are fused to increase its strength while lightening its weight. Most species of bird only possess one ovary rather than two, and no living birds have teeth in their jaw, further reducing body mass.



(a) (b)

(a) Primary feathers are located at the wing tip and provide thrust; secondary feathers are located close to the body and provide lift. (b) Many birds have hollow pneumatic bones, which make flight easier.

Birds possess a system of air sacs branching from their primary airway that divert the path of air so that it passes unidirectionally through the lung, during both inspiration and expiration. Unlike mammalian lungs in which air flows in two directions as it is breathed in and out, air flows continuously through the bird’s lung to provide a more efficient system of gas exchange.

Mammals

Mammals are vertebrates that have hair and mammary glands used to provide nutrition for their young. Certain features of the jaw, skeleton, skin, and internal anatomy are also unique to mammals. The presence of hair is one of the key characteristics of a mammal. Although it is not very extensive in some groups, such as whales, hair has many important functions for mammals. Mammals are endothermic, and hair provides insulation by trapping a layer of air close to the body to retain metabolic heat. Hair also serves as a sensory mechanism through specialized hairs called vibrissae, better known as whiskers. These attach to nerves that transmit touch information, which is particularly useful to nocturnal or burrowing mammals. Hair can also provide protective coloration.

Mammalian skin includes secretory glands with various functions. **Sebaceous glands** produce a lipid mixture called sebum that is secreted onto the hair and skin for water resistance and lubrication. Sebaceous glands are located over most of the body. **Sudoriferous glands** produce sweat and scent, which function in thermoregulation and communication, respectively. **Mammary glands** produce milk that is used to feed newborns. While male monotremes and eutherians possess mammary glands, male marsupials do not.

The skeletal system of mammals possesses unique features that differentiate them from other vertebrates. Most mammals have **heterodont teeth**, meaning they have different types and shapes of teeth that allow them to feed on different kinds of foods. These different types of teeth include the incisors, the canines, premolars, and molars. The first two types are for cutting and tearing, whereas the latter two types are for crushing and grinding. Different groups have different proportions of each type, depending on their diet. Most mammals are also **diphyodonts**, meaning they have two sets of teeth in their lifetime: deciduous or “baby” teeth, and permanent teeth. In other vertebrates, the teeth can be replaced throughout life.

Modern mammals are divided into three broad groups: monotremes, marsupials, and eutherians (or placental mammals). The eutherians, or placental mammals, and the marsupials collectively are called therian mammals, whereas monotremes are called metatherians.

There are three living species of **monotremes**: the platypus and two species of echidnas, or spiny anteaters ([link]). The platypus and one species of echidna are found in Australia, whereas the other species of echidna is found in New Guinea. Monotremes are unique among mammals, as they lay leathery eggs, similar to those of reptiles, rather than giving birth to live young. However, the eggs are retained within the mother’s reproductive tract until they are almost ready to hatch. Once the young hatch, the female begins to secrete milk from pores in a ridge of mammary tissue along the ventral side of her body. Like other mammals, monotremes are endothermic but regulate body temperatures somewhat lower (90 °F, 32 °C) than placental mammals do (98 °F, 37 °C). Like reptiles, monotremes have one posterior opening for urinary, fecal, and reproductive products, rather than three separate openings like placental mammals do. Adult monotremes lack teeth.

The platypus (left), a monotreme, possesses a leathery beak and lays eggs rather than giving birth to live young. An echidna, another monotreme, is shown in the right photo. (credit “echidna”: modification of work by Barry Thomas)

Marsupials are found primarily in Australia and nearby islands, although about 100 species of opossums and a few species of two other families are found in the Americas. Australian marsupials number over 230 species and include the kangaroo, koala, bandicoot, and Tasmanian devil ([link]). Most species of marsupials possess a pouch in which the young reside after birth, receiving milk and continuing to develop. Before birth, marsupials have a less complex placental connection, and the young are born much less developed than in placental mammals.

The Tasmanian devil is one of several marsupials native to Australia. (credit: Wayne McLean)

Eutherians are the most widespread of the mammals, occurring throughout the world. There are several groups of eutherians, including Insectivora, the insect eaters; Edentata, the toothless anteaters; Rodentia, the rodents; Chiroptera, the bats; Cetacea, the aquatic mammals including whales; Carnivora, carnivorous mammals including dogs, cats, and bears; and Primates, which includes humans. **Eutherian mammals** are sometimes called placental mammals, because all species have a complex placenta that connects a fetus to the mother, allowing for gas, fluid, waste, and nutrient exchange. While other mammals may possess a less complex placenta or briefly have a placenta, all eutherians have a complex placenta during gestation.

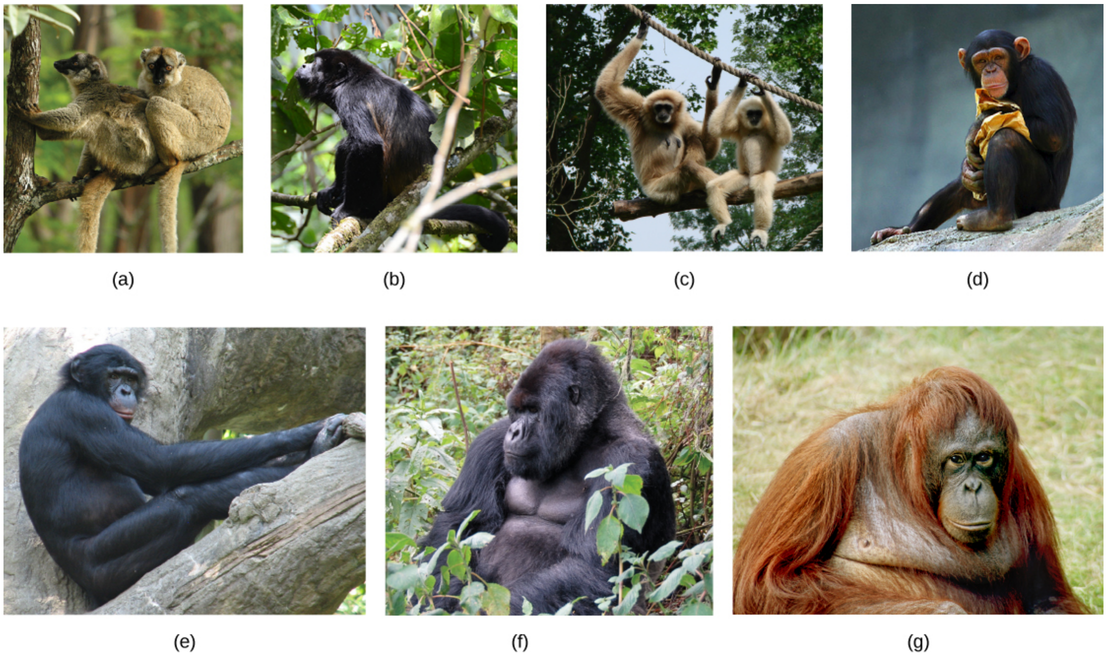
Primates

Order **Primates** of class Mammalia includes lemurs, tarsiers, monkeys, and the apes, which include humans. Non-human primates live primarily in tropical or subtropical regions of South America, Africa, and Asia. They range in size from the mouse lemur at 30 grams (1 ounce) to the mountain

gorilla at 200 kilograms (441 pounds). The characteristics and evolution of primates are of particular interest to us as they allow us to understand the evolution of our own species.

All primate species have adaptations for climbing trees, as they all descended from tree-dwellers, although not all species are arboreal. This arboreal heritage of primates resulted in hands and feet that are adapted for **brachiation**, or climbing and swinging through trees. These adaptations include, but are not limited to 1) a rotating shoulder joint, 2) a big toe that is widely separated from the other toes and thumbs that are widely separated from fingers (except humans), which allow for gripping branches, and 3) **stereoscopic vision**, two overlapping visual fields, which allows for the depth perception necessary to gauge distance. Other characteristics of primates are brains that are larger than those of many other mammals, claws that have been modified into flattened nails, typically only one offspring per pregnancy, and a trend toward holding the body upright.

Order Primates is divided into two groups: prosimians and anthropoids. **Prosimians** include the bush babies of Africa, the lemurs of Madagascar, and the lorises, pottos, and tarsiers of Southeast Asia. **Anthropoids** include monkeys, lesser apes, and great apes ([link]). In general, prosimians tend to be nocturnal, smaller in size than anthropoids, and have relatively smaller brains compared to anthropoids.



Primates can be divided into prosimians, such as the (a) lemur, and anthropoids. Anthropoids include monkeys, such as the (b) howler monkey; lesser apes, such as the (c) gibbon; and great apes, such as the (d) chimpanzee, (e) bonobo, (f) gorilla, and (g) orangutan. (credit a: modification of work by Frank Vassen; credit b: modification of work by Xavi Talleda; credit d: modification of work by Aaron Logan; credit e: modification of work by Trisha Shears; credit f: modification of work by Dave Proffer; credit g: modification of work by Julie Langford)

Section Summary

The earliest vertebrates that diverged from the invertebrate chordates were the jawless fishes. Hagfishes are eel-like scavengers that feed on dead invertebrates and other fishes. Lampreys are characterized by a toothed, funnel-like sucking mouth, and some species are parasitic on other fishes. Gnathostomes include the jawed fishes (cartilaginous and bony fishes) as well as all other tetrapods. Cartilaginous fishes include sharks, rays, skates,

and ghost sharks. Bony fishes can be further divided into ray-finned and lobe-finned fishes.

As tetrapods, most amphibians are characterized by four well-developed limbs, although some species of salamanders and all caecilians are limbless. Amphibians have a moist, permeable skin used for cutaneous respiration. Amphibia can be divided into three clades: salamanders (Urodela), frogs (Anura), and caecilians (Apoda). The life cycle of amphibians consists of two distinct stages: the larval stage and metamorphosis to an adult stage.

The amniotes are distinguished from amphibians by the presence of a terrestrially adapted egg protected by amniotic membranes. The amniotes include reptiles, birds, and mammals. A key adaptation that permitted reptiles to live on land was the development of scaly skin. Reptilia includes four living clades: Crocodilia (crocodiles and alligators), Sphenodontia (tuataras), Squamata (lizards and snakes), and Testudines (turtles).

Birds are endothermic amniotes. Feathers act as insulation and allow for flight. Birds have pneumatic bones that are hollow rather than tissue-filled. Airflow through bird lungs travels in one direction. Birds evolved from dinosaurs.

Mammals have hair and mammary glands. Mammalian skin includes various secretory glands. Mammals are endothermic, like birds. There are three groups of mammals living today: monotremes, marsupials, and eutherians. Monotremes are unique among mammals as they lay eggs, rather than giving birth to live young. Eutherian mammals have a complex placenta.

There are 16 extant (living) orders of eutherian mammals. Humans are most closely related to Primates, all of which have adaptations for climbing trees, although not all species are arboreal. Other characteristics of primates are brains that are larger than those of other mammals, claws that have been modified into flattened nails, and typically one young per pregnancy, stereoscopic vision, and a trend toward holding the body upright. Primates are divided into two groups: prosimians and anthropoids.

Review Questions

Exercise:

Problem:

Members of Chondrichthyes differ from members of Osteichthyes by having a _________.

a. jaw
b. bony skeleton
c. cartilaginous skeleton
d. two sets of paired fins

Solution:

C

Exercise:

Problem: Squamata includes ______.

a. crocodiles and alligators
b. turtles
c. tuataras
d. lizards and snakes

Solution:

D

Exercise:

Problem: Sudoriferous glands produce _________.

a. sweat
b. lipids

c. sebum
d. milk

Solution:

A

Exercise:

Problem: Which of the following is a Monotreme?

a. kangaroo
b. koala
c. bandicoot
d. platypus

Solution:

D

Free Response

Exercise:

Problem:

What can be inferred about the evolution of the cranium and the vertebral column from examining hagfishes and lampreys?

Solution:

Comparison of hagfishes with lampreys shows that the cranium evolved first in early vertebrates, as it is seen in hagfishes, which evolved earlier than lampreys. This was followed by evolution of the vertebral column, a primitive form of which is seen in lampreys and not in hagfishes.

Exercise:

Problem: Explain why frogs are restricted to a moist environment.

Solution:

A moist environment is required as frog eggs lack a shell and dehydrate quickly in dry environments.

Exercise:

Problem: Describe three adaptations that allow for flight in birds.

Solution:

Birds have feathers that streamline the bird body and assist in flight. They also have pneumatic bones that are hollow rather than tissue-filled. Birds are endothermic, which allows for a higher metabolism demanded by flight.

Glossary

Actinopterygii
: ray-finned fishes

amniote
: a clade of animals that possesses an amniotic egg; includes reptiles (including birds) and mammals

Amphibia
: frogs, salamanders, and caecilians

ampulla of Lorenzini
: a sensory organ that allows sharks to detect electromagnetic fields produced by living things

anthropoids

a clade consisting of monkeys, apes, and humans

Anura
frogs

Apoda
caecilians

brachiation
swinging through trees

caecilian
a legless amphibian that belongs to clade Apoda

Chondrichthyes
jawed fishes with paired fins and a skeleton made of cartilage

craniate
a proposed clade of chordates that includes all groups except the tunicates and lancelets

Crocodilia
crocodiles and alligators

cutaneous respiration
gas exchange through the skin

diphyodont
refers to the possession of two sets of teeth in a lifetime

down feather
feather specialized for insulation

eutherian mammal
a mammal with a complex placenta, which connects a fetus to the mother; sometimes called placental mammals

frog
a tail-less amphibian that belongs to clade Anura

gnathostome
: a jawed fish

hagfish
: an eel-like jawless fish that lives on the ocean floor and is a scavenger

heterodont teeth
: different types of teeth modified by different purposes

lamprey
: a jawless fish characterized by a toothed, funnel-like, sucking mouth

lateral line
: the sense organ that runs the length of a fish’s body, used to detect vibration in the water

mammal
: one of the groups of endothermic vertebrates that possess hair and mammary glands

mammary gland
: in female mammals, a gland that produces milk for newborns

marsupial
: one of the groups of mammals that includes the kangaroo, koala, bandicoot, Tasmanian devil, and several other species; young develop within a pouch

monotreme
: an egg-laying mammal

Myxini
: hagfishes

Osteichthyes
: bony fishes

ostracoderm
: one of the earliest jawless fishes covered in bone

Petromyzontidae
: the clade of lampreys

pneumatic bone
: an air-filled bone

Primates
: includes lemurs, tarsiers, monkeys, apes, and humans

prosimians
: a group of primates that includes bush babies of Africa, lemurs of Madagascar, and lorises, pottos, and tarsiers of southeast Asia

salamander
: a tailed amphibian that belongs to the clade Urodela

Sarcopterygii
: lobe-finned fishes

sebaceous gland
: in mammals, a skin gland that produce a lipid mixture called sebum

Sphenodontia
: the reptilian clade that includes the tuataras

Squamata
: the reptilian clade of lizards and snakes

stereoscopic vision
: two overlapping fields of vision from the eyes that produces depth perception

sudoriferous gland
: a gland in mammals that produces sweat and scent molecules

swim bladder
: in fishes, a gas filled organ that helps to control the buoyancy of the fish

tadpole
the larval stage of a frog

Testudines
turtles

Urodela
salamanders

The Plant Body
By the end of this section, you will be able to:

- Describe the shoot organ system and the root organ system
- Distinguish between meristematic tissue and permanent tissue
- Identify and describe the three regions where plant growth occurs
- Summarize the roles of dermal tissue, vascular tissue, and ground tissue
- Compare simple plant tissue with complex plant tissue

Like animals, plants contain cells with organelles in which specific metabolic activities take place. Unlike animals, however, plants use energy from sunlight to form sugars during photosynthesis. In addition, plant cells have cell walls, plastids, and a large central vacuole: structures that are not found in animal cells. Each of these cellular structures plays a specific role in plant structure and function.

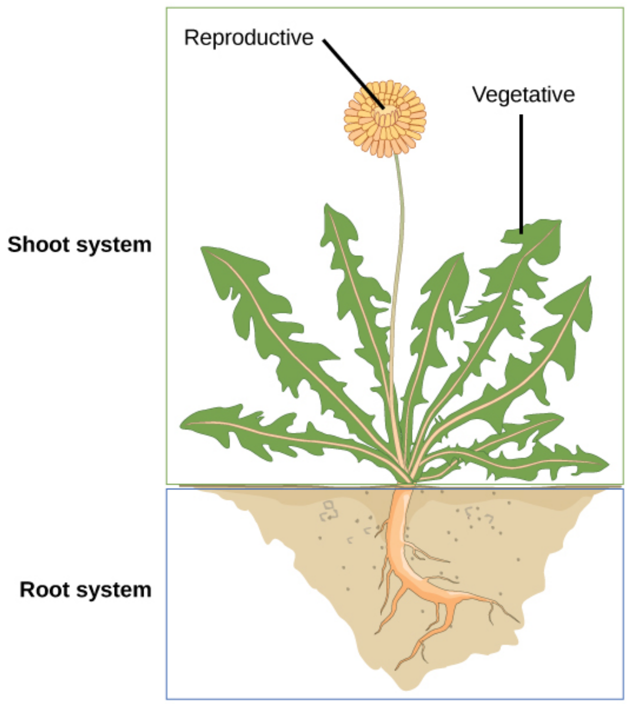
Note:
Link to Learning

Watch *Botany Without Borders*, a video produced by the Botanical Society of America about the importance of plants.

Plant Organ Systems

In plants, just as in animals, similar cells working together form a tissue. When different types of tissues work together to perform a unique function, they form an organ; organs working together form organ systems. Vascular

plants have two distinct organ systems: a shoot system, and a root system. The **shoot system** consists of two portions: the vegetative (non-reproductive) parts of the plant, such as the leaves and the stems, and the reproductive parts of the plant, which include flowers and fruits. The shoot system generally grows above ground, where it absorbs the light needed for photosynthesis. The **root system**, which supports the plants and absorbs water and minerals, is usually underground. [link] shows the organ systems of a typical plant.



The shoot system of a plant consists of leaves, stems, flowers, and fruits. The root system anchors the plant while absorbing water and minerals from the soil.

Plant Tissues

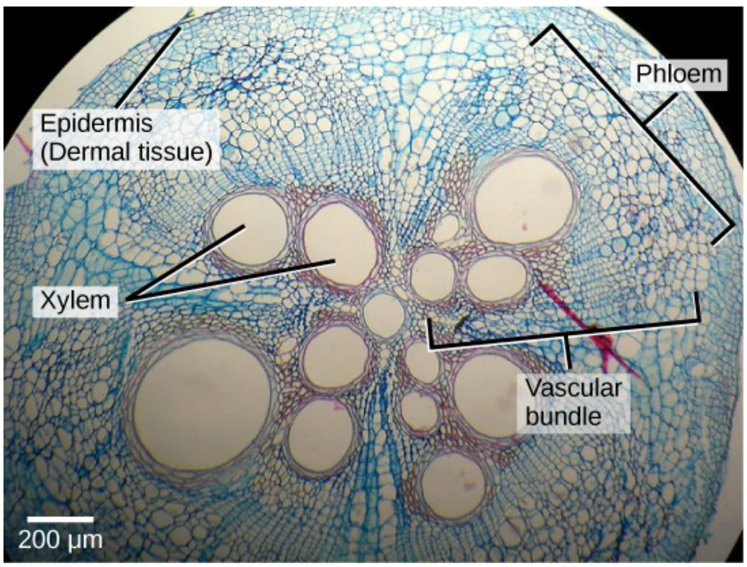
Plants are multicellular eukaryotes with tissue systems made of various cell types that carry out specific functions. Plant tissue systems fall into one of two general types: meristematic tissue, and permanent (or non-meristematic) tissue. Cells of the meristematic tissue are found in **meristems**, which are plant regions of continuous cell division and growth. **Meristematic tissue** cells are either undifferentiated or incompletely differentiated, and they continue to divide and contribute to the growth of the plant. In contrast, **permanent tissue** consists of plant cells that are no longer actively dividing.

Meristematic tissues consist of three types, based on their location in the plant. **Apical meristems** contain meristematic tissue located at the tips of stems and roots, which enable a plant to extend in length. **Lateral meristems** facilitate growth in thickness or girth in a maturing plant. **Intercalary meristems** occur only in monocots, at the bases of leaf blades and at nodes (the areas where leaves attach to a stem). This tissue enables the monocot leaf blade to increase in length from the leaf base; for example, it allows lawn grass leaves to elongate even after repeated mowing.

Meristems produce cells that quickly differentiate, or specialize, and become permanent tissue. Such cells take on specific roles and lose their ability to divide further. They differentiate into three main types: dermal, vascular, and ground tissue. **Dermal tissue** covers and protects the plant, and **vascular tissue** transports water, minerals, and sugars to different parts of the plant. **Ground tissue** serves as a site for photosynthesis, provides a supporting matrix for the vascular tissue, and helps to store water and sugars.

Secondary tissues are either simple (composed of similar cell types) or complex (composed of different cell types). Dermal tissue, for example, is a simple tissue that covers the outer surface of the plant and controls gas exchange. Vascular tissue is an example of a complex tissue, and is made of two specialized conducting tissues: xylem and phloem. Xylem tissue transports water and nutrients from the roots to different parts of the plant, and includes three different cell types: vessel elements and tracheids (both of which conduct water), and xylem parenchyma. Phloem tissue, which transports organic compounds from the site of photosynthesis to other parts

of the plant, consists of four different cell types: sieve cells (which conduct photosynthates), companion cells, phloem parenchyma, and phloem fibers. Unlike xylem conducting cells, phloem conducting cells are alive at maturity. The xylem and phloem always lie adjacent to each other ([link]). In stems, the xylem and the phloem form a structure called a **vascular bundle**; in roots, this is termed the **vascular stele** or **vascular cylinder**.



This light micrograph shows a cross section of a squash (*Cucurbita maxima*) stem. Each teardrop-shaped vascular bundle consists of large xylem vessels toward the inside and smaller phloem cells toward the outside. Xylem cells, which transport water and nutrients from the roots to the rest of the plant, are dead at functional maturity. Phloem cells, which transport sugars and other organic compounds from photosynthetic tissue to the rest of the plant, are living. The vascular bundles are encased in ground tissue and surrounded by dermal tissue. (credit: modification of work by "

(biophotos)"/Flickr; scale-bar data from Matt Russell)

Section Summary

A vascular plant consists of two organ systems: the shoot system and the root system. The shoot system includes the aboveground vegetative portions (stems and leaves) and reproductive parts (flowers and fruits). The root system supports the plant and is usually underground. A plant is composed of two main types of tissue: meristematic tissue and permanent tissue. Meristematic tissue consists of actively dividing cells found in root and shoot tips. As growth occurs, meristematic tissue differentiates into permanent tissue, which is categorized as either simple or complex. Simple tissues are made up of similar cell types; examples include dermal tissue and ground tissue. Dermal tissue provides the outer covering of the plant. Ground tissue is responsible for photosynthesis; it also supports vascular tissue and may store water and sugars. Complex tissues are made up of different cell types. Vascular tissue, for example, is made up of xylem and phloem cells.

Review Questions

Exercise:

Problem:

Plant regions of continuous growth are made up of _________.

a. dermal tissue
b. vascular tissue
c. meristematic tissue
d. permanent tissue

Solution:

C

Exercise:

Problem: Which of the following is the major site of photosynthesis?

a. apical meristem
b. ground tissue
c. xylem cells
d. phloem cells

Solution:

B

Free Response

Exercise:

Problem:

What type of meristem is found only in monocots, such as lawn grasses? Explain how this type of meristematic tissue is beneficial in lawn grasses that are mowed each week.

Solution:

Lawn grasses and other monocots have an intercalary meristem, which is a region of meristematic tissue at the base of the leaf blade. This is beneficial to the plant because it can continue to grow even when the tip of the plant is removed by grazing or mowing.

Exercise:

Problem:

Which plant part is responsible for transporting water, minerals, and sugars to different parts of the plant? Name the two types of tissue that make up this overall tissue, and explain the role of each.

Solution:

Vascular tissue transports water, minerals, and sugars throughout the plant. Vascular tissue is made up of xylem tissue and phloem tissue. Xylem tissue transports water and nutrients from the roots upward. Phloem tissue carries sugars from the sites of photosynthesis to the rest of the plant.

Glossary

apical meristem
: meristematic tissue located at the tips of stems and roots; enables a plant to extend in length

dermal tissue
: protective plant tissue covering the outermost part of the plant; controls gas exchange

ground tissue
: plant tissue involved in photosynthesis; provides support, and stores water and sugars

intercalary meristem
: meristematic tissue located at nodes and the bases of leaf blades; found only in monocots

lateral meristem
: meristematic tissue that enables a plant to increase in thickness or girth

meristematic tissue

tissue containing cells that constantly divide; contributes to plant growth

meristem
: plant region of continuous growth

permanent tissue
: plant tissue composed of cells that are no longer actively dividing

root system
: belowground portion of the plant that supports the plant and absorbs water and minerals

shoot system
: aboveground portion of the plant; consists of non-reproductive plant parts, such as leaves and stems, and reproductive parts, such as flowers and fruits

vascular bundle
: strands of stem tissue made up of xylem and phloem

vascular stele
: strands of root tissue made up of xylem and phloem

vascular tissue
: tissue made up of xylem and phloem that transports food and water throughout the plant

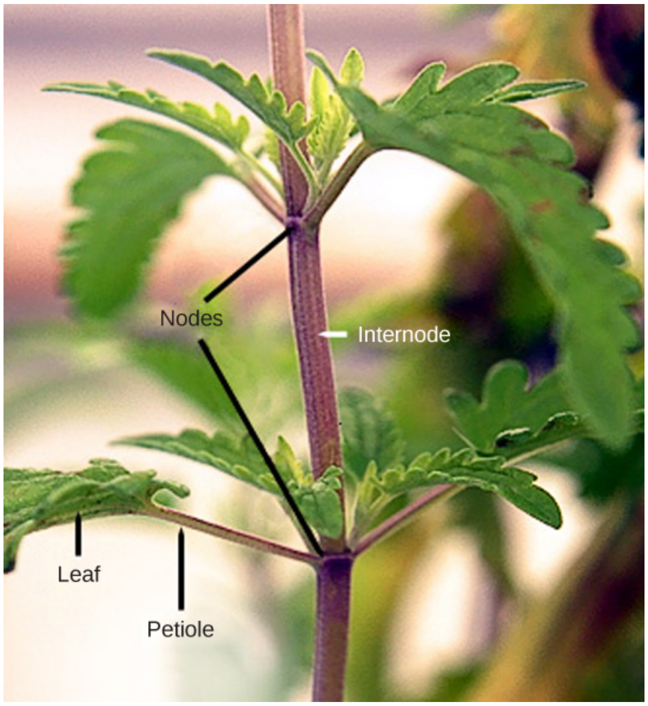
Stems

By the end of this section, you will be able to:

- Describe the main function and basic structure of stems
- Compare and contrast the roles of dermal tissue, vascular tissue, and ground tissue
- Distinguish between primary growth and secondary growth in stems
- Summarize the origin of annual rings
- List and describe examples of modified stems

Stems are a part of the shoot system of a plant. They may range in length from a few millimeters to hundreds of meters, and also vary in diameter, depending on the plant type. Stems are usually above ground, although the stems of some plants, such as the potato, also grow underground. Stems may be herbaceous (soft) or woody in nature. Their main function is to provide support to the plant, holding leaves, flowers and buds; in some cases, stems also store food for the plant. A stem may be unbranched, like that of a palm tree, or it may be highly branched, like that of a magnolia tree. The stem of the plant connects the roots to the leaves, helping to transport absorbed water and minerals to different parts of the plant. It also helps to transport the products of photosynthesis, namely sugars, from the leaves to the rest of the plant.

Plant stems, whether above or below ground, are characterized by the presence of nodes and internodes ([link]). **Nodes** are points of attachment for leaves, aerial roots, and flowers. The stem region between two nodes is called an **internode**. The stalk that extends from the stem to the base of the leaf is the petiole. An **axillary bud** is usually found in the axil—the area between the base of a leaf and the stem—where it can give rise to a branch or a flower. The apex (tip) of the shoot contains the apical meristem within the **apical bud**.



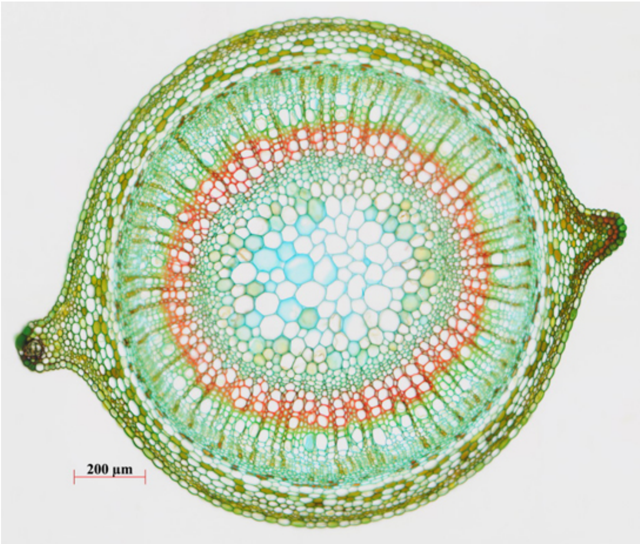
Leaves are attached to the plant stem at areas called nodes. An internode is the stem region between two nodes. The petiole is the stalk connecting the leaf to the stem. The leaves just above the nodes arose from axillary buds.

Stem Anatomy

The stem and other plant organs arise from the ground tissue, and are primarily made up of simple tissues formed from three types of cells: parenchyma, collenchyma, and sclerenchyma cells.

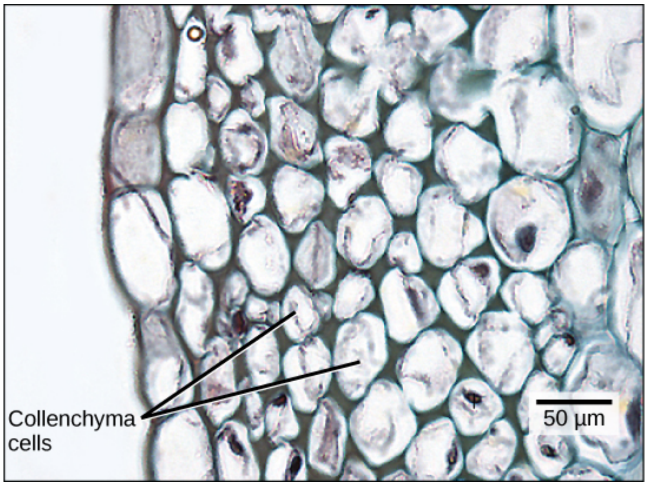
Parenchyma cells are the most common plant cells ([link]). They are found in the stem, the root, the inside of the leaf, and the pulp of the fruit. Parenchyma cells are responsible for metabolic functions, such as

photosynthesis, and they help repair and heal wounds. Some parenchyma cells also store starch.



The stem of common St John's Wort (*Hypericum perforatum*) is shown in cross section in this light micrograph. The central pith (greenish-blue, in the center) and peripheral cortex (narrow zone 3–5 cells thick just inside the epidermis) are composed of parenchyma cells. Vascular tissue composed of xylem (red) and phloem tissue (green, between the xylem and cortex) surrounds the pith. (credit: Rolf-Dieter Mueller)

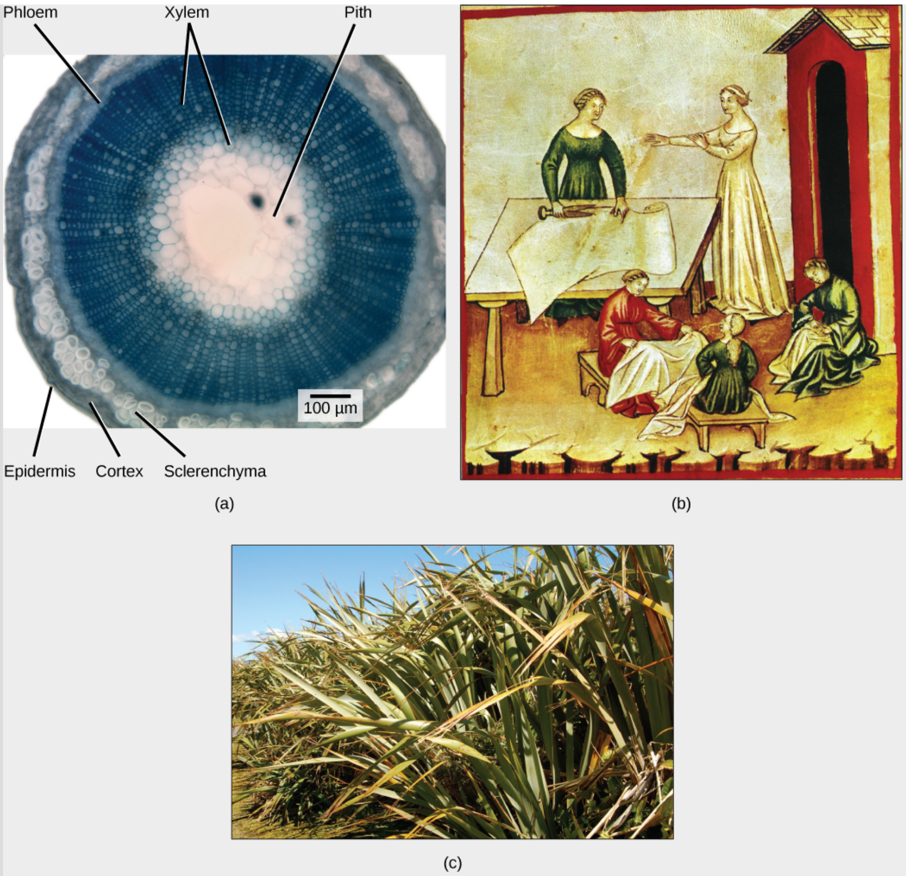
Collenchyma cells are elongated cells with unevenly thickened walls ([link]). They provide structural support, mainly to the stem and leaves. These cells are alive at maturity and are usually found below the epidermis. The “strings” of a celery stalk are an example of collenchyma cells.



Collenchyma cell walls are uneven in thickness, as seen in this light micrograph. They provide support to plant structures. (credit: modification of work by Carl Szczerski; scale-bar data from Matt Russell)

Sclerenchyma cells also provide support to the plant, but unlike collenchyma cells, many of them are dead at maturity. There are two types of sclerenchyma cells: fibers and sclereids. Both types have secondary cell walls that are thickened with deposits of lignin, an organic compound that is a key component of wood. Fibers are long, slender cells; sclereids are smaller-sized. Sclereids give pears their gritty texture. Humans use sclerenchyma fibers to make linen and rope ([link]).

Note:
Art Connection



The central pith and outer cortex of the (a) flax stem are made up of parenchyma cells. Inside the cortex is a layer of sclerenchyma cells, which make up the fibers in flax rope and clothing. Humans have grown and harvested flax for thousands of years. In (b) this drawing, fourteenth-century women prepare linen. The (c) flax plant is grown and harvested for its fibers, which are used to weave linen, and for its seeds, which are the source of linseed oil. (credit a: modification of work by Emmanuel Boutet based on original work by Ryan R. MacKenzie; credit c: modification of work by Brian Dearth; scale-bar data from Matt Russell)

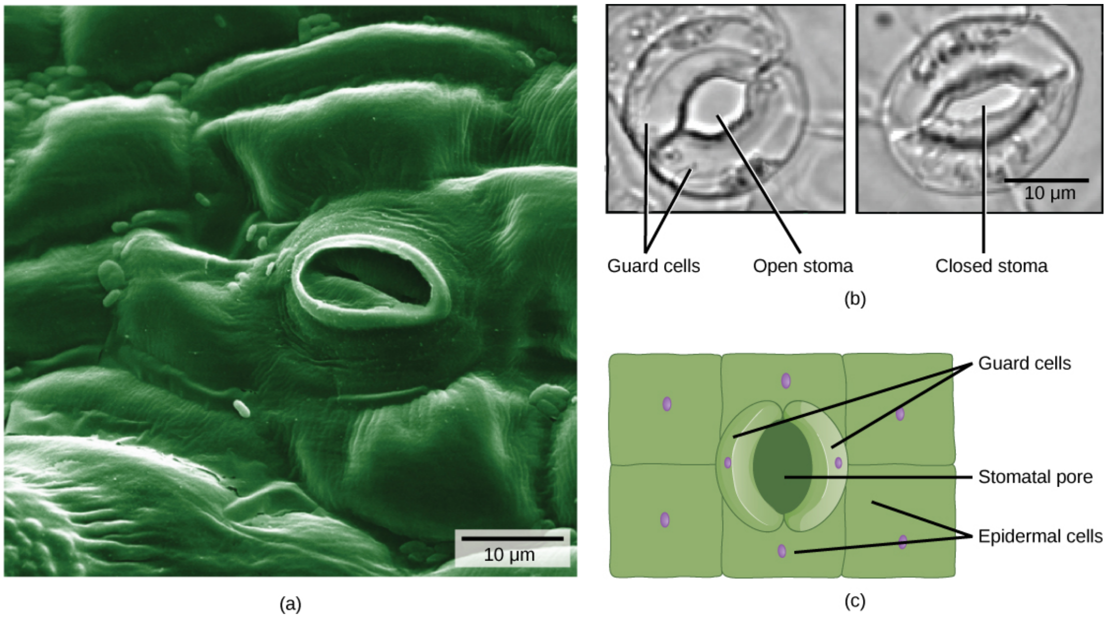
Which layers of the stem are made of parenchyma cells?

a. cortex and pith
b. phloem
c. sclerenchyma
d. xylem

Like the rest of the plant, the stem has three tissue systems: dermal, vascular, and ground tissue. Each is distinguished by characteristic cell types that perform specific tasks necessary for the plant’s growth and survival.

Dermal Tissue

The dermal tissue of the stem consists primarily of **epidermis**, a single layer of cells covering and protecting the underlying tissue. Woody plants have a tough, waterproof outer layer of cork cells commonly known as **bark**, which further protects the plant from damage. Epidermal cells are the most numerous and least differentiated of the cells in the epidermis. The epidermis of a leaf also contains openings known as stomata, through which the exchange of gases takes place ([link]). Two cells, known as **guard cells**, surround each leaf stoma, controlling its opening and closing and thus regulating the uptake of carbon dioxide and the release of oxygen and water vapor. **Trichomes** are hair-like structures on the epidermal surface. They help to reduce **transpiration** (the loss of water by aboveground plant parts), increase solar reflectance, and store compounds that defend the leaves against predation by herbivores.

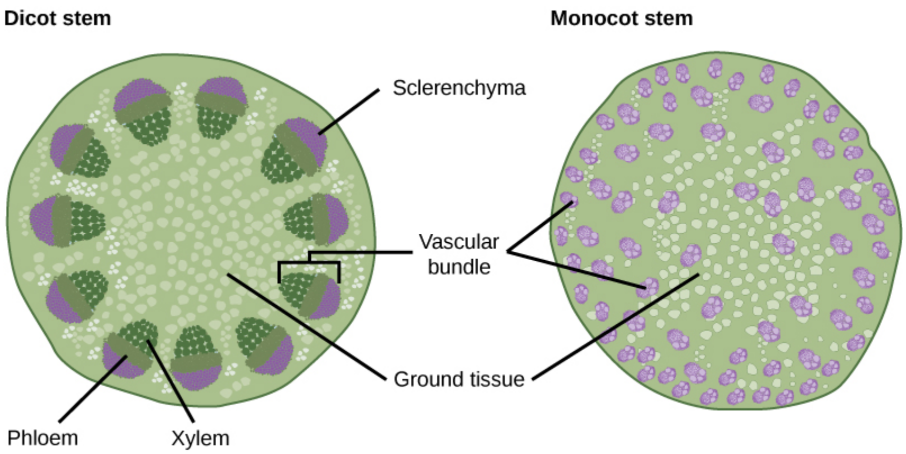


Openings called stomata (singular: stoma) allow a plant to take up carbon dioxide and release oxygen and water vapor. The (a) colorized scanning-electron micrograph shows a closed stoma of a dicot. Each stoma is flanked by two guard cells that regulate its (b) opening and closing. The (c) guard cells sit within the layer of epidermal cells (credit a: modification of work by Louisa Howard, Rippel Electron Microscope Facility, Dartmouth College; credit b: modification of work by June Kwak, University of Maryland; scale-bar data from Matt Russell)

Vascular Tissue

The xylem and phloem that make up the vascular tissue of the stem are arranged in distinct strands called vascular bundles, which run up and down the length of the stem. When the stem is viewed in cross section, the vascular bundles of dicot stems are arranged in a ring. In plants with stems that live for more than one year, the individual bundles grow together and

produce the characteristic growth rings. In monocot stems, the vascular bundles are randomly scattered throughout the ground tissue ([link]).



In (a) dicot stems, vascular bundles are arranged around the periphery of the ground tissue. The xylem tissue is located toward the interior of the vascular bundle, and phloem is located toward the exterior. Sclerenchyma fibers cap the vascular bundles. In (b) monocot stems, vascular bundles composed of xylem and phloem tissues are scattered throughout the ground tissue.

Xylem tissue has three types of cells: xylem parenchyma, tracheids, and vessel elements. The latter two types conduct water and are dead at maturity. **Tracheids** are xylem cells with thick secondary cell walls that are lignified. Water moves from one tracheid to another through regions on the side walls known as pits, where secondary walls are absent. **Vessel elements** are xylem cells with thinner walls; they are shorter than tracheids. Each vessel element is connected to the next by means of a perforation plate at the end walls of the element. Water moves through the perforation plates to travel up the plant.

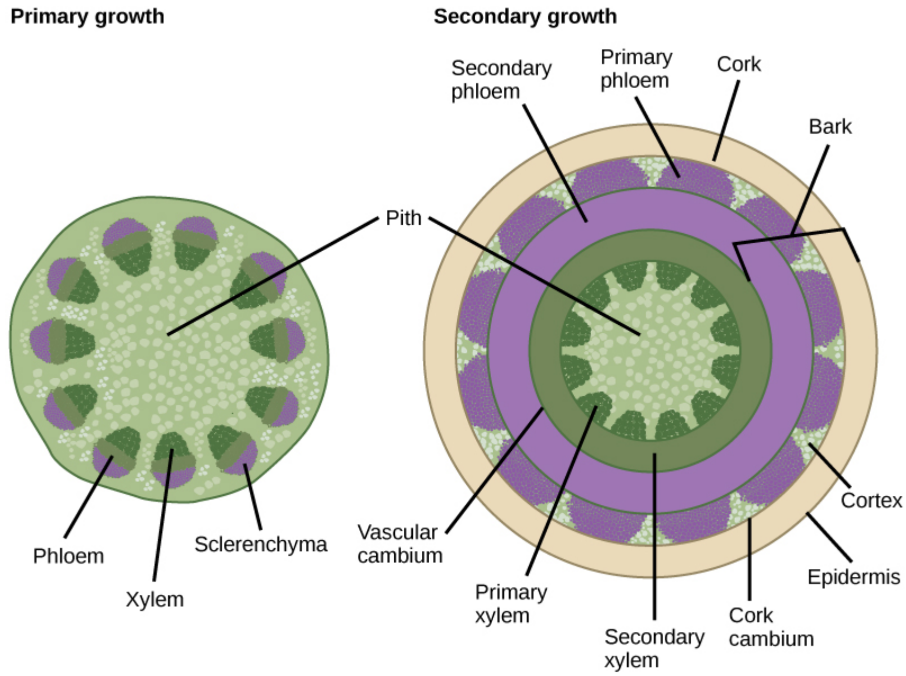
Phloem tissue is composed of sieve-tube cells, companion cells, phloem parenchyma, and phloem fibers. A series of **sieve-tube cells** (also called sieve-tube elements) are arranged end to end to make up a long sieve tube, which transports organic substances such as sugars and amino acids. The sugars flow from one sieve-tube cell to the next through perforated sieve plates, which are found at the end junctions between two cells. Although still alive at maturity, the nucleus and other cell components of the sieve-tube cells have disintegrated. **Companion cells** are found alongside the sieve-tube cells, providing them with metabolic support. The companion cells contain more ribosomes and mitochondria than the sieve-tube cells, which lack some cellular organelles.

Ground Tissue

Ground tissue is mostly made up of parenchyma cells, but may also contain collenchyma and sclerenchyma cells that help support the stem. The ground tissue towards the interior of the vascular tissue in a stem or root is known as **pith**, while the layer of tissue between the vascular tissue and the epidermis is known as the **cortex**.

Growth in Stems

Growth in plants occurs as the stems and roots lengthen. Some plants, especially those that are woody, also increase in thickness during their life span. The increase in length of the shoot and the root is referred to as **primary growth**, and is the result of cell division in the shoot apical meristem. **Secondary growth** is characterized by an increase in thickness or girth of the plant, and is caused by cell division in the lateral meristem. [link] shows the areas of primary and secondary growth in a plant. Herbaceous plants mostly undergo primary growth, with hardly any secondary growth or increase in thickness. Secondary growth or “wood” is noticeable in woody plants; it occurs in some dicots, but occurs very rarely in monocots.



In woody plants, primary growth is followed by secondary growth, which allows the plant stem to increase in thickness or girth. Secondary vascular tissue is added as the plant grows, as well as a cork layer. The bark of a tree extends from the vascular cambium to the epidermis.

Some plant parts, such as stems and roots, continue to grow throughout a plant's life: a phenomenon called indeterminate growth. Other plant parts, such as leaves and flowers, exhibit determinate growth, which ceases when a plant part reaches a particular size.

Primary Growth

Most primary growth occurs at the apices, or tips, of stems and roots. Primary growth is a result of rapidly dividing cells in the apical meristems at the shoot tip and root tip. Subsequent cell elongation also contributes to

primary growth. The growth of shoots and roots during primary growth enables plants to continuously seek water (roots) or sunlight (shoots).

The influence of the apical bud on overall plant growth is known as apical dominance, which diminishes the growth of axillary buds that form along the sides of branches and stems. Most coniferous trees exhibit strong apical dominance, thus producing the typical conical Christmas tree shape. If the apical bud is removed, then the axillary buds will start forming lateral branches. Gardeners make use of this fact when they prune plants by cutting off the tops of branches, thus encouraging the axillary buds to grow out, giving the plant a bushy shape.

Note:
Link to Learning

Watch this BBC Nature video showing how time-lapse photography captures plant growth at high speed.

Secondary Growth

The increase in stem thickness that results from secondary growth is due to the activity of the lateral meristems, which are lacking in herbaceous plants. Lateral meristems include the vascular cambium and, in woody plants, the cork cambium (see [link]). The vascular cambium is located just outside the primary xylem and to the interior of the primary phloem. The cells of the vascular cambium divide and form secondary xylem (tracheids and vessel elements) to the inside, and secondary phloem (sieve elements and companion cells) to the outside. The thickening of the stem that occurs in

secondary growth is due to the formation of secondary phloem and secondary xylem by the vascular cambium, plus the action of cork cambium, which forms the tough outermost layer of the stem. The cells of the secondary xylem contain lignin, which provides hardiness and strength.

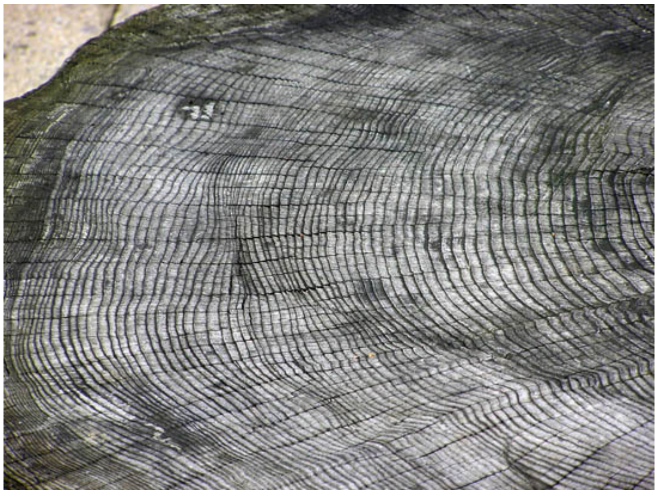
In woody plants, cork cambium is the outermost lateral meristem. It produces cork cells (bark) containing a waxy substance known as suberin that can repel water. The bark protects the plant against physical damage and helps reduce water loss. The cork cambium also produces a layer of cells known as phelloderm, which grows inward from the cambium. The cork cambium, cork cells, and phelloderm are collectively termed the **periderm**. The periderm substitutes for the epidermis in mature plants. In some plants, the periderm has many openings, known as **lenticels**, which allow the interior cells to exchange gases with the outside atmosphere ([link]). This supplies oxygen to the living and metabolically active cells of the cortex, xylem and phloem.

Lenticels on the bark of this cherry tree enable the woody stem to exchange gases with

the surrounding atmosphere.
(credit: Roger Griffith)

Annual Rings

The activity of the vascular cambium gives rise to annual growth rings. During the spring growing season, cells of the secondary xylem have a large internal diameter and their primary cell walls are not extensively thickened. This is known as early wood, or spring wood. During the fall season, the secondary xylem develops thickened cell walls, forming late wood, or autumn wood, which is denser than early wood. This alternation of early and late wood is due largely to a seasonal decrease in the number of vessel elements and a seasonal increase in the number of tracheids. It results in the formation of an annual ring, which can be seen as a circular ring in the cross section of the stem ([link]). An examination of the number of annual rings and their nature (such as their size and cell wall thickness) can reveal the age of the tree and the prevailing climatic conditions during each season.



The rate of wood growth increases

in summer and decreases in winter, producing a characteristic ring for each year of growth. Seasonal changes in weather patterns can also affect the growth rate—note how the rings vary in thickness. (credit: Adrian Pingstone)

Stem Modifications

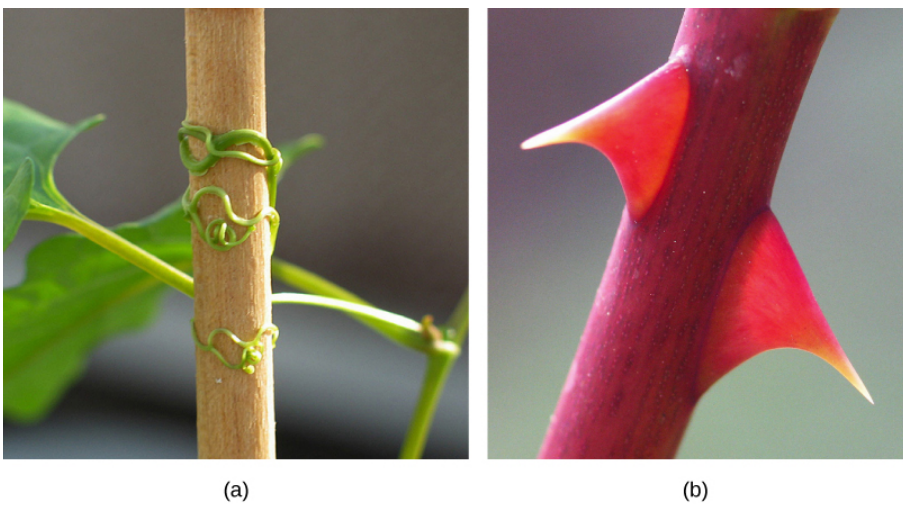
Some plant species have modified stems that are especially suited to a particular habitat and environment ([link]). A **rhizome** is a modified stem that grows horizontally underground and has nodes and internodes. Vertical shoots may arise from the buds on the rhizome of some plants, such as ginger and ferns. **Corms** are similar to rhizomes, except they are more rounded and fleshy (such as in gladiolus). Corms contain stored food that enables some plants to survive the winter. **Stolons** are stems that run almost parallel to the ground, or just below the surface, and can give rise to new plants at the nodes. **Runners** are a type of stolon that runs above the ground and produces new clone plants at nodes at varying intervals: strawberries are an example. **Tubers** are modified stems that may store starch, as seen in the potato (*Solanum* sp.). Tubers arise as swollen ends of stolons, and contain many adventitious or unusual buds (familiar to us as the “eyes” on potatoes). A **bulb**, which functions as an underground storage unit, is a modification of a stem that has the appearance of enlarged fleshy leaves emerging from the stem or surrounding the base of the stem, as seen in the iris.

Stem modifications enable plants to thrive in a variety of environments. Shown are (a) ginger (*Zingiber officinale*) rhizomes, (b) a carrion flower (*Amorphophallus titanum*) corm (c) Rhodes grass (*Chloris gayana*) stolons, (d) strawberry (*Fragaria ananassa*) runners, (e) potato (*Solanum tuberosum*) tubers, and (f) red onion (*Allium*) bulbs. (credit a: modification of work by Maja Dumat; credit c: modification of work by Harry Rose; credit d: modification of work by Rebecca Siegel; credit e: modification of work by Scott Bauer, USDA ARS; credit f: modification of work by Stephen Ausmus, USDA ARS)

Note:
Link to Learning

Watch botanist Wendy Hodgson, of Desert Botanical Garden in Phoenix, Arizona, explain how agave plants were cultivated for food hundreds of years ago in the Arizona desert in this video: *Finding the Roots of an Ancient Crop.*

Some aerial modifications of stems are tendrils and thorns ([link]). **Tendrils** are slender, twining strands that enable a plant (like a vine or pumpkin) to seek support by climbing on other surfaces. **Thorns** are modified branches appearing as sharp outgrowths that protect the plant; common examples include roses, Osage orange and devil's walking stick.



Found in southeastern United States, (a) buckwheat vine (*Brunnichia ovata*) is a weedy plant that climbs with the aid of tendrils. This one is shown climbing up a wooden stake. (b) Thorns are modified branches. (credit a: modification of work by Christopher Meloche, USDA ARS; credit b: modification of work by "macrophile"/Flickr)

Section Summary

The stem of a plant bears the leaves, flowers, and fruits. Stems are characterized by the presence of nodes (the points of attachment for leaves or branches) and internodes (regions between nodes).

Plant organs are made up of simple and complex tissues. The stem has three tissue systems: dermal, vascular, and ground tissue. Dermal tissue is the outer covering of the plant. It contains epidermal cells, stomata, guard cells, and trichomes. Vascular tissue is made up of xylem and phloem tissues and conducts water, minerals, and photosynthetic products. Ground tissue is responsible for photosynthesis and support and is composed of parenchyma, collenchyma, and sclerenchyma cells.

Primary growth occurs at the tips of roots and shoots, causing an increase in length. Woody plants may also exhibit secondary growth, or increase in thickness. In woody plants, especially trees, annual rings may form as growth slows at the end of each season. Some plant species have modified stems that help to store food, propagate new plants, or discourage predators. Rhizomes, corms, stolons, runners, tubers, bulbs, tendrils, and thorns are examples of modified stems.

Art Connections

Exercise:

Problem:

[link] Which layers of the stem are made of parenchyma cells?

A. cortex and pith
B. epidermis
C. sclerenchyma
D. epidermis and cortex.

Solution:

[link] A and B. The cortex, pith, and epidermis are made of parenchyma cells.

Review Questions

Exercise:

Problem:

Stem regions at which leaves are attached are called ________.

a. trichomes
b. lenticels
c. nodes
d. internodes

Solution:

C

Exercise:

Problem:

Which of the following cell types forms most of the inside of a plant?

a. meristem cells
b. collenchyma cells
c. sclerenchyma cells
d. parenchyma cells

Solution:

D

Exercise:

Problem:

Tracheids, vessel elements, sieve-tube cells, and companion cells are components of _________.

a. vascular tissue
b. meristematic tissue
c. ground tissue
d. dermal tissue

Solution:

A

Exercise:

Problem:

The primary growth of a plant is due to the action of the _________.

a. lateral meristem
b. vascular cambium
c. apical meristem
d. cork cambium

Solution:

C

Exercise:

Problem: Which of the following is an example of secondary growth?

a. increase in length
b. increase in thickness or girth
c. increase in root hairs
d. increase in leaf number

Solution:

B

Exercise:

Problem: Secondary growth in stems is usually seen in ________.

a. monocots
b. dicots
c. both monocots and dicots
d. neither monocots nor dicots

Solution:

B

Free Response

Exercise:

Problem:

Describe the roles played by stomata and guard cells. What would happen to a plant if these cells did not function correctly?

Solution:

Stomata allow gases to enter and exit the plant. Guard cells regulate the opening and closing of stomata. If these cells did not function correctly, a plant could not get the carbon dioxide needed for photosynthesis, nor could it release the oxygen produced by photosynthesis.

Exercise:

Problem:

Compare the structure and function of xylem to that of phloem.

Solution:

Xylem is made up tracheids and vessel elements, which are cells that transport water and dissolved minerals and that are dead at maturity. Phloem is made up of sieve-tube cells and companion cells, which transport carbohydrates and are alive at maturity.

Exercise:

Problem: Explain the role of the cork cambium in woody plants.

Solution:

In woody plants, the cork cambium is the outermost lateral meristem; it produces new cells towards the interior, which enables the plant to increase in girth. The cork cambium also produces cork cells towards the exterior, which protect the plant from physical damage while reducing water loss.

Exercise:

Problem: What is the function of lenticels?

Solution:

In woody stems, lenticels allow internal cells to exchange gases with the outside atmosphere.

Exercise:

Problem:

Besides the age of a tree, what additional information can annual rings reveal?

Solution:

Annual rings can also indicate the climate conditions that prevailed during each growing season.

Exercise:

Problem:

Give two examples of modified stems and explain how each example benefits the plant.

Solution:

Answers will vary. Rhizomes, stolons, and runners can give rise to new plants. Corms, tubers, and bulbs can also produce new plants and can store food. Tendrils help a plant to climb, while thorns discourage herbivores.

Glossary

apical bud
: bud formed at the tip of the shoot

axillary bud
: bud located in the axil: the stem area where the petiole connects to the stem

bark
: tough, waterproof, outer epidermal layer of cork cells

bulb
: modified underground stem that consists of a large bud surrounded by numerous leaf scales

collenchyma cell
: elongated plant cell with unevenly thickened walls; provides structural support to the stem and leaves

companion cell
phloem cell that is connected to sieve-tube cells; has large amounts of ribosomes and mitochondrion

corm
rounded, fleshy underground stem that contains stored food

cortex
ground tissue found between the vascular tissue and the epidermis in a stem or root

epidermis
single layer of cells found in plant dermal tissue; covers and protects underlying tissue

guard cells
paired cells on either side of a stoma that control stomatal opening and thereby regulate the movement of gases and water vapor

internode
region between nodes on the stem

lenticel
opening on the surface of mature woody stems that facilitates gas exchange

node
point along the stem at which leaves, flowers, or aerial roots originate

parenchyma cell
most common type of plant cell; found in the stem, root, leaf, and in fruit pulp; site of photosynthesis and starch storage

periderm
outermost covering of woody stems; consists of the cork cambium, cork cells, and the phelloderm

pith

ground tissue found towards the interior of the vascular tissue in a stem or root

primary growth
: growth resulting in an increase in length of the stem and the root; caused by cell division in the shoot or root apical meristem

rhizome
: modified underground stem that grows horizontally to the soil surface and has nodes and internodes

runner
: stolon that runs above the ground and produces new clone plants at nodes

sclerenchyma cell
: plant cell that has thick secondary walls and provides structural support; usually dead at maturity

secondary growth
: growth resulting in an increase in thickness or girth; caused by the lateral meristem and cork cambium

sieve-tube cell
: phloem cell arranged end to end to form a sieve tube that transports organic substances such as sugars and amino acids

stolon
: modified stem that runs parallel to the ground and can give rise to new plants at the nodes

tendril
: modified stem consisting of slender, twining strands used for support or climbing

thorn
: modified stem branch appearing as a sharp outgrowth that protects the plant

tracheid
xylem cell with thick secondary walls that helps transport water

trichome
hair-like structure on the epidermal surface

tuber
modified underground stem adapted for starch storage; has many adventitious buds

vessel element
xylem cell that is shorter than a tracheid and has thinner walls

Roots

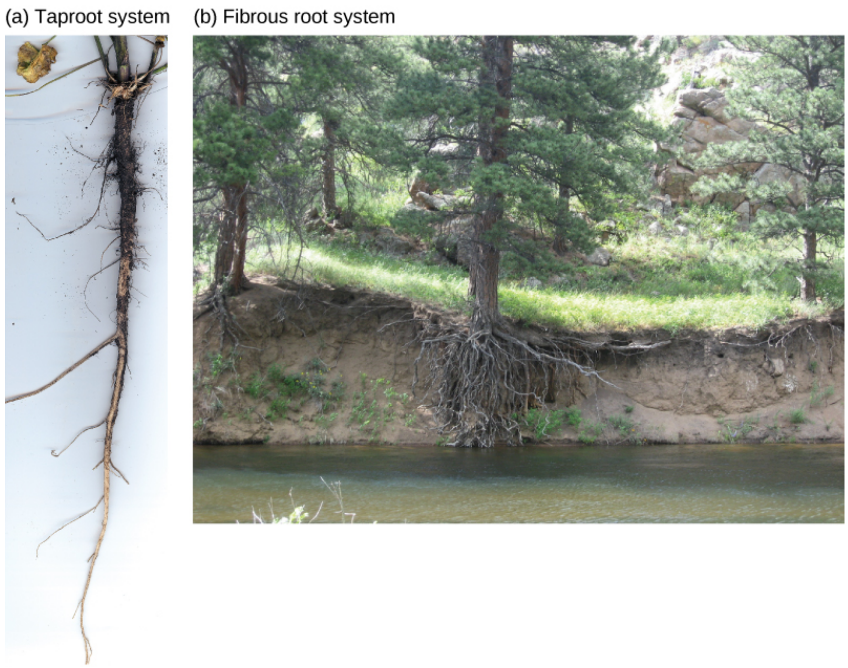
By the end of this section, you will be able to:

- Identify the two types of root systems
- Describe the three zones of the root tip and summarize the role of each zone in root growth
- Describe the structure of the root
- List and describe examples of modified roots

The roots of seed plants have three major functions: anchoring the plant to the soil, absorbing water and minerals and transporting them upwards, and storing the products of photosynthesis. Some roots are modified to absorb moisture and exchange gases. Most roots are underground. Some plants, however, also have **adventitious roots**, which emerge above the ground from the shoot.

Types of Root Systems

Root systems are mainly of two types ([link]). Dicots have a tap root system, while monocots have a fibrous root system. A **tap root system** has a main root that grows down vertically, and from which many smaller lateral roots arise. Dandelions are a good example; their tap roots usually break off when trying to pull these weeds, and they can regrow another shoot from the remaining root). A tap root system penetrates deep into the soil. In contrast, a **fibrous root system** is located closer to the soil surface, and forms a dense network of roots that also helps prevent soil erosion (lawn grasses are a good example, as are wheat, rice, and corn). Some plants have a combination of tap roots and fibrous roots. Plants that grow in dry areas often have deep root systems, whereas plants growing in areas with abundant water are likely to have shallower root systems.

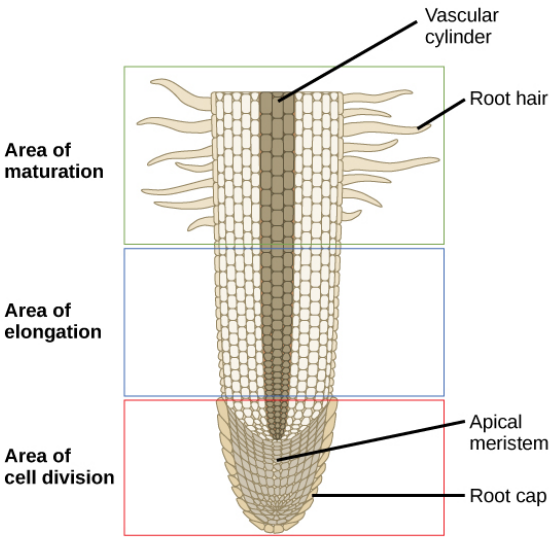


(a) Tap root systems have a main root that grows down, while (b) fibrous root systems consist of many small roots. (credit b: modification of work by “Austen Squarepants”/Flickr)

Root Growth and Anatomy

Root growth begins with seed germination. When the plant embryo emerges from the seed, the radicle of the embryo forms the root system. The tip of the root is protected by the **root cap**, a structure exclusive to roots and unlike any other plant structure. The root cap is continuously replaced because it gets damaged easily as the root pushes through soil. The root tip can be divided into three zones: a zone of cell division, a zone of elongation, and a zone of maturation and differentiation ([link]). The zone of cell division is closest to the root tip; it is made up of the actively dividing cells of the root meristem. The zone of elongation is where the newly formed cells increase in length, thereby lengthening the root. Beginning at the first root hair is the zone of cell maturation where the root

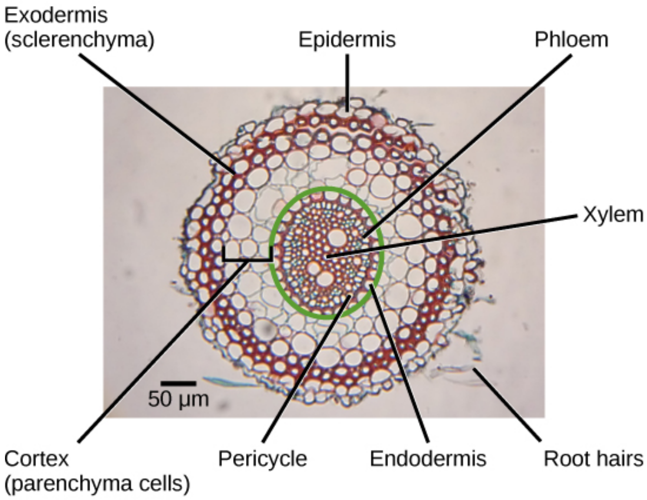
cells begin to differentiate into special cell types. All three zones are in the first centimeter or so of the root tip.



A longitudinal view of the root reveals the zones of cell division, elongation, and maturation. Cell division occurs in the apical meristem.

The root has an outer layer of cells called the epidermis, which surrounds areas of ground tissue and vascular tissue. The epidermis provides protection and helps in absorption. **Root hairs**, which are extensions of root epidermal cells, increase the surface area of the root, greatly contributing to the absorption of water and minerals.

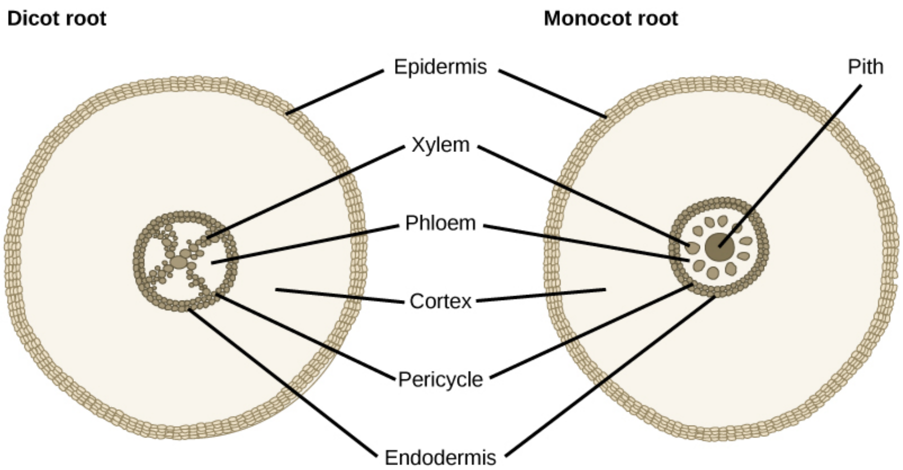
Inside the root, the ground tissue forms two regions: the cortex and the pith ([link]). Compared to stems, roots have lots of cortex and little pith. Both regions include cells that store photosynthetic products. The cortex is between the epidermis and the vascular tissue, whereas the pith lies between the vascular tissue and the center of the root.



Staining reveals different cell types in this light micrograph of a wheat (*Triticum*) root cross section. Sclerenchyma cells of the exodermis and xylem cells stain red, and phloem cells stain blue. Other cell types stain black. The stele, or vascular tissue, is the area inside endodermis (indicated by a green ring). Root hairs are visible outside the epidermis. (credit: scale-bar data from Matt Russell)

The vascular tissue in the root is arranged in the inner portion of the root, which is called the **stele** ([link]). A layer of cells known as the **endodermis** separates the stele from the ground tissue in the outer portion of the root. The endodermis is exclusive to roots, and serves as a checkpoint for materials entering the root's vascular system. A waxy substance called suberin is present on the walls of the endodermal cells. This waxy region, known as the **Casparian strip**, forces water and solutes to cross the plasma membranes of endodermal cells instead of slipping between the cells. This ensures that only materials required by the root pass through the endodermis, while toxic substances and pathogens are generally excluded.

The outermost cell layer of the root's vascular tissue is the **pericycle**, an area that can give rise to lateral roots. In dicot roots, the xylem and phloem of the stele are arranged alternately in an X shape, whereas in monocot roots, the vascular tissue is arranged in a ring around the pith.



In (left) typical dicots, the vascular tissue forms an X shape in the center of the root. In (right) typical monocots, the phloem cells and the larger xylem cells form a characteristic ring around the central pith.

Root Modifications

Root structures may be modified for specific purposes. For example, some roots are bulbous and store starch. Aerial roots and prop roots are two forms of aboveground roots that provide additional support to anchor the plant. Tap roots, such as carrots, turnips, and beets, are examples of roots that are modified for food storage ([link]).

Many vegetables are modified roots.

Epiphytic roots enable a plant to grow on another plant. For example, the epiphytic roots of orchids develop a spongy tissue to absorb moisture. The banyan tree (*Ficus* sp.) begins as an epiphyte, germinating in the branches of a host tree; aerial roots develop from the branches and eventually reach the ground, providing additional support ([link]). In screwpine (*Pandanus* sp.), a palm-like tree that grows in sandy tropical soils, aboveground prop roots develop from the nodes to provide additional support.

(a) (b)

The (a) banyan tree, also known as the strangler fig, begins life as an epiphyte in a host tree. Aerial roots extend to the ground and support the growing plant,

which eventually strangles the host tree. The (b) screwpine develops aboveground roots that help support the plant in sandy soils. (credit a: modification of work by "psyberartist"/Flickr; credit b: modification of work by David Eikhoff)

Section Summary

Roots help to anchor a plant, absorb water and minerals, and serve as storage sites for food. Taproots and fibrous roots are the two main types of root systems. In a taproot system, a main root grows vertically downward with a few lateral roots. Fibrous root systems arise at the base of the stem, where a cluster of roots forms a dense network that is shallower than a taproot. The growing root tip is protected by a root cap. The root tip has three main zones: a zone of cell division (cells are actively dividing), a zone of elongation (cells increase in length), and a zone of maturation (cells differentiate to form different kinds of cells). Root vascular tissue conducts water, minerals, and sugars. In some habitats, the roots of certain plants may be modified to form aerial roots or epiphytic roots.

Review Questions

Exercise:

Problem:

Roots that enable a plant to grow on another plant are called ________.

a. epiphytic roots
b. prop roots
c. adventitious roots
d. aerial roots

Solution:

A

Exercise:

Problem: The _________ forces selective uptake of minerals in the root.

a. pericycle
b. epidermis
c. endodermis
d. root cap

Solution:

C

Exercise:

Problem:

Newly-formed root cells begin to form different cell types in the _________.

a. zone of elongation
b. zone of maturation
c. root meristem
d. zone of cell division

Solution:

B

Free Response

Exercise:

Problem:

Compare a tap root system with a fibrous root system. For each type, name a plant that provides a food in the human diet. Which type of root system is found in monocots? Which type of root system is found in dicots?

Solution:

A tap root system has a single main root that grows down. A fibrous root system forms a dense network of roots that is closer to the soil surface. An example of a tap root system is a carrot. Grasses such as wheat, rice, and corn are examples of fibrous root systems. Fibrous root systems are found in monocots; tap root systems are found in dicots.

Exercise:

Problem: What might happen to a root if the pericycle disappeared?

Solution:

The root would not be able to produce lateral roots.

Glossary

adventitious root
: aboveground root that arises from a plant part other than the radicle of the plant embryo

Casparian strip
: waxy coating that forces water to cross endodermal plasma membranes before entering the vascular cylinder, instead of moving between endodermal cells

endodermis

layer of cells in the root that forms a selective barrier between the ground tissue and the vascular tissue, allowing water and minerals to enter the root while excluding toxins and pathogens

fibrous root system
: type of root system in which the roots arise from the base of the stem in a cluster, forming a dense network of roots; found in monocots

pericycle
: outer boundary of the stele from which lateral roots can arise

root cap
: protective cells covering the tip of the growing root

root hair
: hair-like structure that is an extension of epidermal cells; increases the root surface area and aids in absorption of water and minerals

stele
: inner portion of the root containing the vascular tissue; surrounded by the endodermis

tap root system
: type of root system with a main root that grows vertically with few lateral roots; found in dicots

Leaves
By the end of this section, you will be able to:

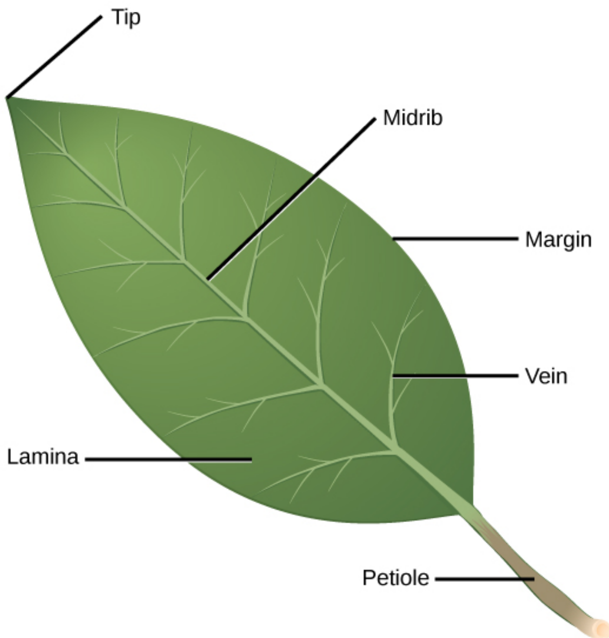
- Identify the parts of a typical leaf
- Describe the internal structure and function of a leaf
- Compare and contrast simple leaves and compound leaves
- List and describe examples of modified leaves

Leaves are the main sites for photosynthesis: the process by which plants synthesize food. Most leaves are usually green, due to the presence of chlorophyll in the leaf cells. However, some leaves may have different colors, caused by other plant pigments that mask the green chlorophyll.

The thickness, shape, and size of leaves are adapted to the environment. Each variation helps a plant species maximize its chances of survival in a particular habitat. Usually, the leaves of plants growing in tropical rainforests have larger surface areas than those of plants growing in deserts or very cold conditions, which are likely to have a smaller surface area to minimize water loss.

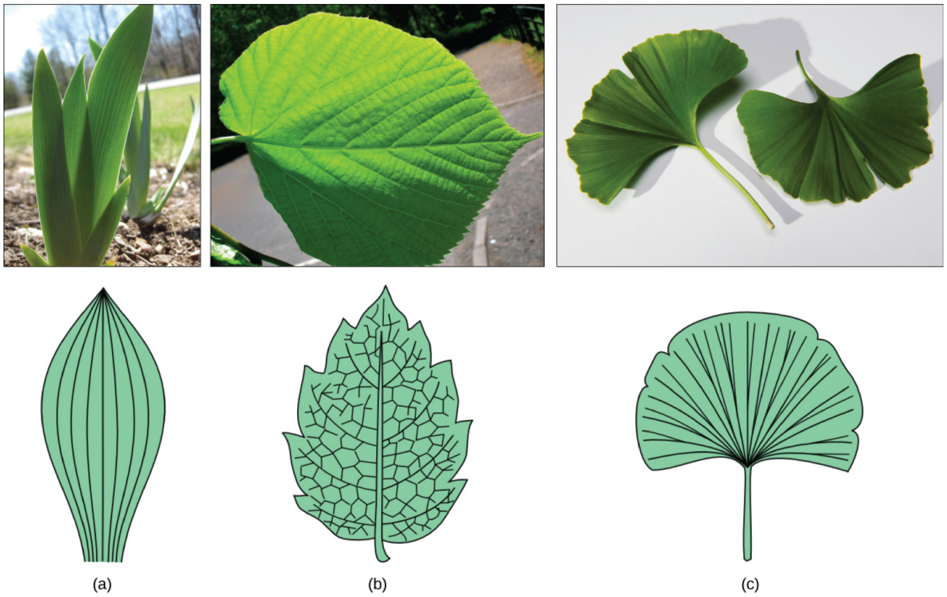
Structure of a Typical Leaf

Each leaf typically has a leaf blade called the **lamina**, which is also the widest part of the leaf. Some leaves are attached to the plant stem by a **petiole**. Leaves that do not have a petiole and are directly attached to the plant stem are called **sessile** leaves. Small green appendages usually found at the base of the petiole are known as **stipules**. Most leaves have a midrib, which travels the length of the leaf and branches to each side to produce veins of vascular tissue. The edge of the leaf is called the margin. [link] shows the structure of a typical eudicot leaf.



Deceptively simple in appearance, a leaf is a highly efficient structure.

Within each leaf, the vascular tissue forms veins. The arrangement of veins in a leaf is called the **venation** pattern. Monocots and dicots differ in their patterns of venation ([link]). Monocots have parallel venation; the veins run in straight lines across the length of the leaf without converging at a point. In dicots, however, the veins of the leaf have a net-like appearance, forming a pattern known as reticulate venation. One extant plant, the *Ginkgo biloba*, has dichotomous venation where the veins fork.



(a) Tulip (*Tulipa*), a monocot, has leaves with parallel venation. The netlike venation in this (b) linden (*Tilia cordata*) leaf distinguishes it as a dicot. The (c) *Ginkgo biloba* tree has dichotomous venation. (credit a photo: modification of work by “Drewboy64”/Wikimedia Commons; credit b photo: modification of work by Roger Griffith; credit c photo: modification of work by "geishaboy500"/Flickr; credit abc illustrations: modification of work by Agnieszka Kwiecień)

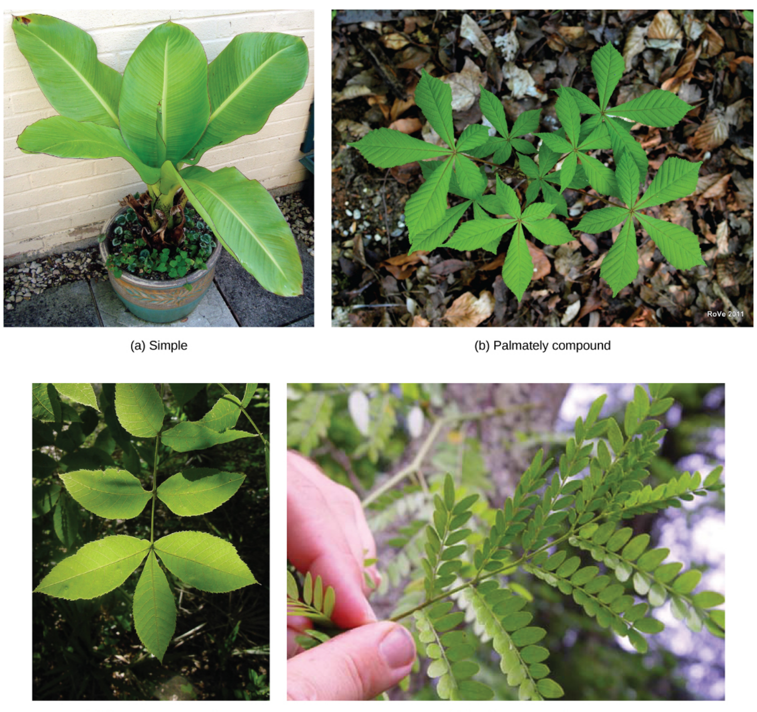
Leaf Arrangement

The arrangement of leaves on a stem is known as **phyllotaxy**. The number and placement of a plant’s leaves will vary depending on the species, with each species exhibiting a characteristic leaf arrangement. Leaves are classified as either alternate, spiral, or opposite. Plants that have only one leaf per node have leaves that are said to be either alternate—meaning the leaves alternate on each side of the stem in a flat plane—or spiral, meaning the leaves are arrayed in a spiral along the stem. In an opposite leaf arrangement, two leaves arise at the same point, with the leaves connecting

opposite each other along the branch. If there are three or more leaves connected at a node, the leaf arrangement is classified as **whorled**.

Leaf Form

Leaves may be simple or compound ([link]). In a **simple leaf**, the blade is either completely undivided—as in the banana leaf—or it has lobes, but the separation does not reach the midrib, as in the maple leaf. In a **compound leaf**, the leaf blade is completely divided, forming leaflets, as in the locust tree. Each leaflet may have its own stalk, but is attached to the rachis. A **palmately compound leaf** resembles the palm of a hand, with leaflets radiating outwards from one point Examples include the leaves of poison ivy, the buckeye tree, or the familiar houseplant *Schefflera* sp. (common name "umbrella plant"). **Pinnately compound leaves** take their name from their feather-like appearance; the leaflets are arranged along the midrib, as in rose leaves (*Rosa* sp.), or the leaves of hickory, pecan, ash, or walnut trees.

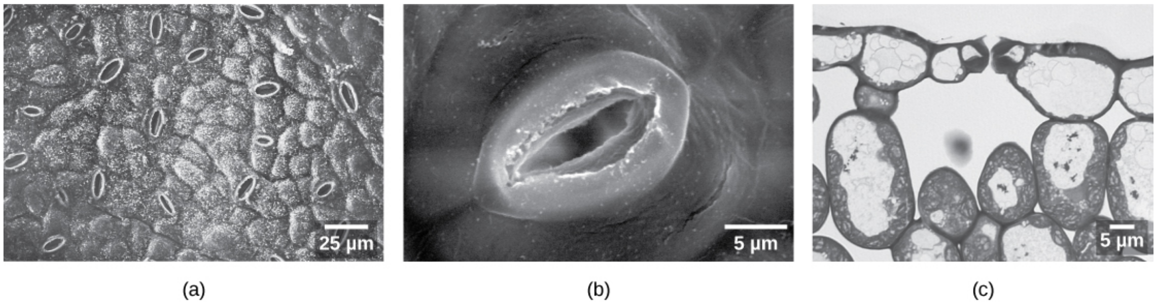


(a) Simple (b) Palmately compound (c) Pinnately compound (d) Doubly compound

Leaves may be simple or compound. In simple leaves, the lamina is continuous. The (a) banana plant (*Musa* sp.) has simple leaves. In compound leaves, the lamina is separated into leaflets. Compound leaves may be palmate or pinnate. In (b) palmately compound leaves, such as those of the horse chestnut (*Aesculus hippocastanum*), the leaflets branch from the petiole. In (c) pinnately compound leaves, the leaflets branch from the midrib, as on a scrub hickory (*Carya floridana*). The (d) honey locust has double compound leaves, in which leaflets branch from the veins. (credit a: modification of work by "BazzaDaRambler"/Flickr; credit b: modification of work by Roberto Verzo; credit c: modification of work by Eric Dion; credit d: modification of work by Valerie Lykes)

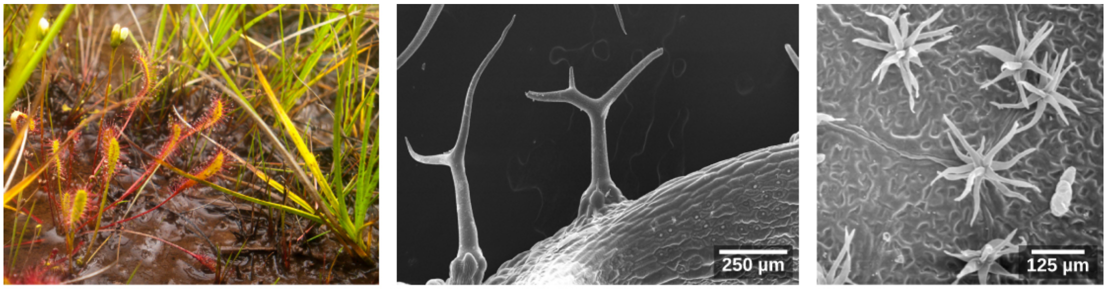
Leaf Structure and Function

The outermost layer of the leaf is the epidermis; it is present on both sides of the leaf and is called the upper and lower epidermis, respectively. Botanists call the upper side the adaxial surface (or adaxis) and the lower side the abaxial surface (or abaxis). The epidermis helps in the regulation of gas exchange. It contains stomata ([link]): openings through which the exchange of gases takes place. Two guard cells surround each stoma, regulating its opening and closing.



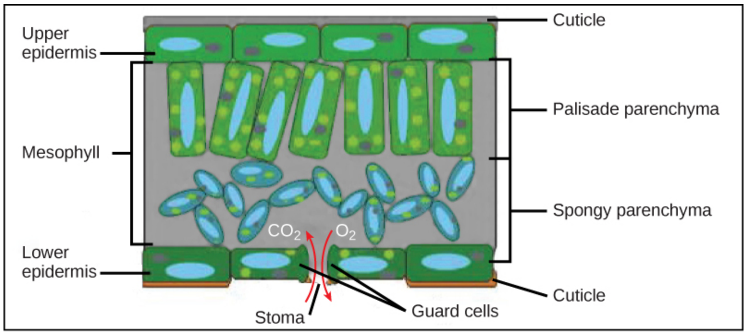
Visualized at 500x with a scanning electron microscope, several stomata are clearly visible on (a) the surface of this sumac (*Rhus glabra*) leaf. At 5,000x magnification, the guard cells of (b) a single stoma from lyre-leaved sand cress (*Arabidopsis lyrata)* have the appearance of lips that surround the opening. In this (c) light micrograph cross-section of an *A. lyrata* leaf, the guard cell pair is visible along with the large, sub-stomatal air space in the leaf. (credit: modification of work by Robert R. Wise; part c scale-bar data from Matt Russell)

The epidermis is usually one cell layer thick; however, in plants that grow in very hot or very cold conditions, the epidermis may be several layers thick to protect against excessive water loss from transpiration. A waxy layer known as the **cuticle** covers the leaves of all plant species. The cuticle reduces the rate of water loss from the leaf surface. Other leaves may have small hairs (trichomes) on the leaf surface. Trichomes help to deter herbivory by restricting insect movements, or by storing toxic or bad-tasting compounds; they can also reduce the rate of transpiration by blocking air flow across the leaf surface ([link]).

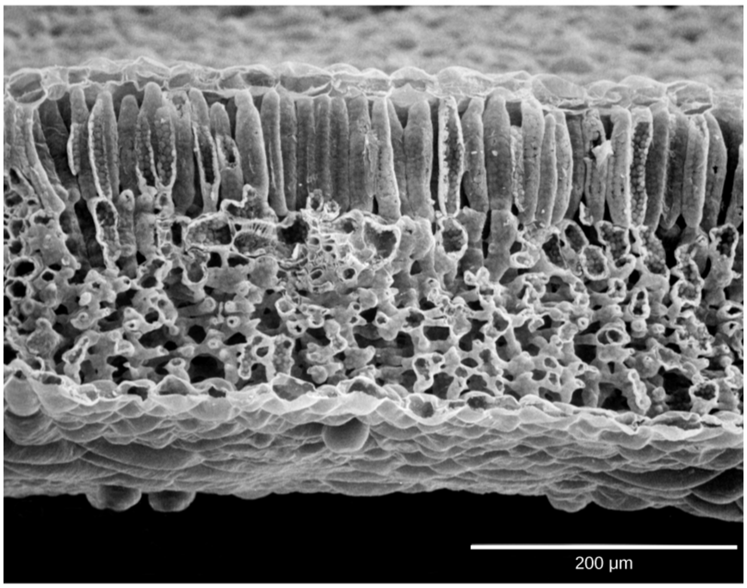


Trichomes give leaves a fuzzy appearance as in this (a) sundew (*Drosera* sp.). Leaf trichomes include (b) branched trichomes on the leaf of *Arabidopsis lyrata* and (c) multibranched trichomes on a mature *Quercus marilandica* leaf. (credit a: John Freeland; credit b, c: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Below the epidermis of dicot leaves are layers of cells known as the mesophyll, or “middle leaf.” The mesophyll of most leaves typically contains two arrangements of parenchyma cells: the palisade parenchyma and spongy parenchyma ([link]). The palisade parenchyma (also called the palisade mesophyll) has column-shaped, tightly packed cells, and may be present in one, two, or three layers. Below the palisade parenchyma are loosely arranged cells of an irregular shape. These are the cells of the spongy parenchyma (or spongy mesophyll). The air space found between the spongy parenchyma cells allows gaseous exchange between the leaf and the outside atmosphere through the stomata. In aquatic plants, the intercellular spaces in the spongy parenchyma help the leaf float. Both layers of the mesophyll contain many chloroplasts. Guard cells are the only epidermal cells to contain chloroplasts.



(a)

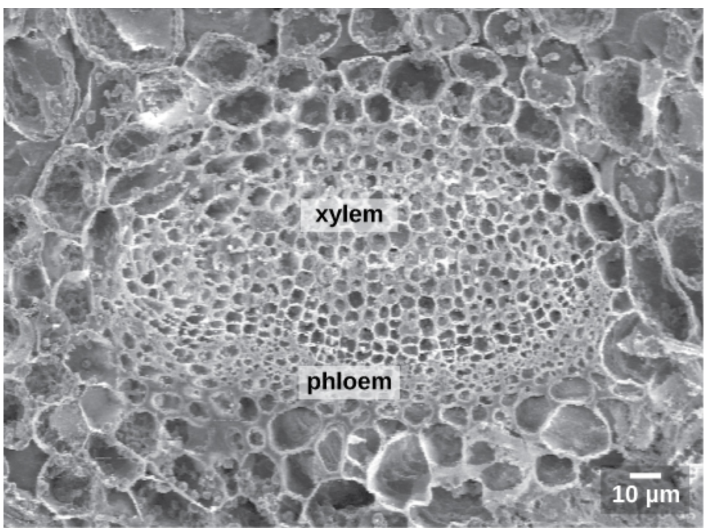


(b)

In the (a) leaf drawing, the central mesophyll is sandwiched between an upper and lower epidermis. The mesophyll has two layers: an upper palisade layer comprised of tightly packed, columnar cells, and a lower spongy layer, comprised of loosely packed, irregularly shaped cells. Stomata on the leaf underside allow gas exchange. A waxy cuticle covers all aerial surfaces of land plants to minimize water loss. These leaf layers are clearly visible in the (b) scanning electron micrograph. The numerous small bumps

in the palisade parenchyma cells are chloroplasts. Chloroplasts are also present in the spongy parenchyma, but are not as obvious. The bumps protruding from the lower surface of the leave are glandular trichomes, which differ in structure from the stalked trichomes in [link]. (credit b: modification of work by Robert R. Wise)

Like the stem, the leaf contains vascular bundles composed of xylem and phloem ([link]). The xylem consists of tracheids and vessels, which transport water and minerals to the leaves. The phloem transports the photosynthetic products from the leaf to the other parts of the plant. A single vascular bundle, no matter how large or small, always contains both xylem and phloem tissues.



This scanning electron micrograph shows xylem and phloem in the leaf vascular bundle from the lyre-leaved sand cress (*Arabidopsis lyrata)*.

(credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Leaf Adaptations

Coniferous plant species that thrive in cold environments, like spruce, fir, and pine, have leaves that are reduced in size and needle-like in appearance. These needle-like leaves have sunken stomata and a smaller surface area: two attributes that aid in reducing water loss. In hot climates, plants such as cacti have leaves that are reduced to spines, which in combination with their succulent stems, help to conserve water. Many aquatic plants have leaves with wide lamina that can float on the surface of the water, and a thick waxy cuticle on the leaf surface that repels water.

Note:
Link to Learning

Watch “The Pale Pitcher Plant” episode of the video series *Plants Are Cool, Too*, a Botanical Society of America video about a carnivorous plant species found in Louisiana.

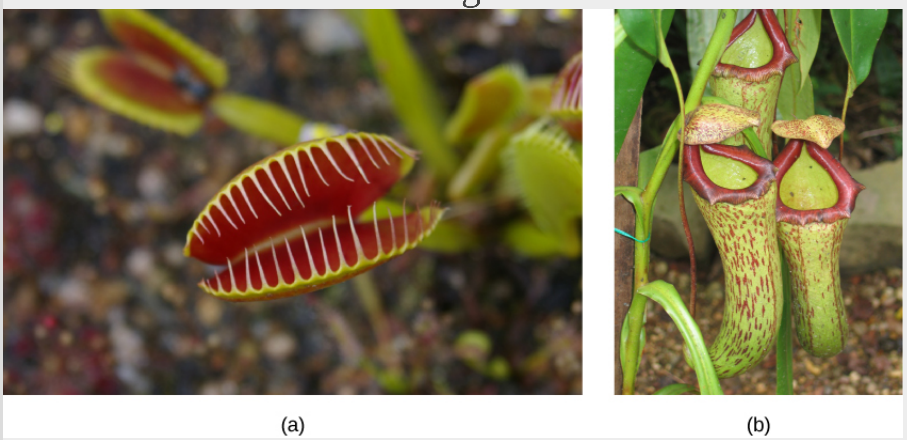
Note:
Evolution Connection
Plant Adaptations in Resource-Deficient Environments

Roots, stems, and leaves are structured to ensure that a plant can obtain the required sunlight, water, soil nutrients, and oxygen resources. Some remarkable adaptations have evolved to enable plant species to thrive in less than ideal habitats, where one or more of these resources is in short supply.

In tropical rainforests, light is often scarce, since many trees and plants grow close together and block much of the sunlight from reaching the forest floor. Many tropical plant species have exceptionally broad leaves to maximize the capture of sunlight. Other species are epiphytes: plants that grow on other plants that serve as a physical support. Such plants are able to grow high up in the canopy atop the branches of other trees, where sunlight is more plentiful. Epiphytes live on rain and minerals collected in the branches and leaves of the supporting plant. Bromeliads (members of the pineapple family), ferns, and orchids are examples of tropical epiphytes ([link]). Many epiphytes have specialized tissues that enable them to efficiently capture and store water.

One of the most well known bromeliads is Spanish moss (*Tillandsia usneoides*), seen here in an oak tree. (credit: Kristine Paulus)

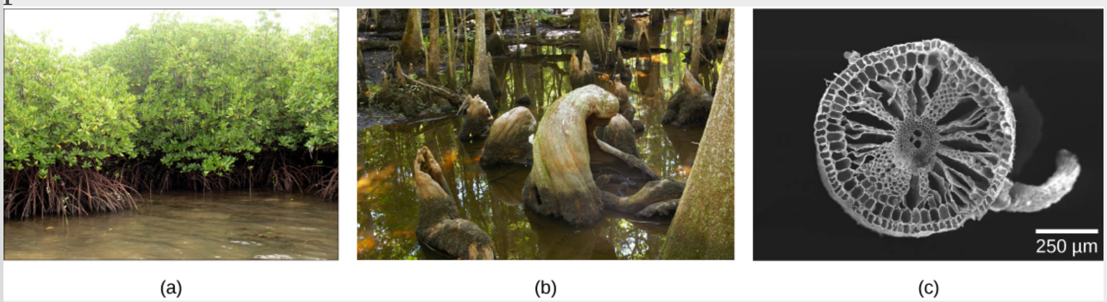
Some plants have special adaptations that help them to survive in nutrient-poor environments. Carnivorous plants, such as the Venus flytrap and the pitcher plant ([link]), grow in bogs where the soil is low in nitrogen. In these plants, leaves are modified to capture insects. The insect-capturing leaves may have evolved to provide these plants with a supplementary source of much-needed nitrogen.



The (a) Venus flytrap has modified leaves that can capture insects. When an unlucky insect touches the trigger hairs inside the leaf, the trap suddenly closes. The opening of the (b) pitcher plant is lined with a slippery wax. Insects crawling on the lip slip and fall into a pool of water in the bottom of the pitcher, where they are digested by bacteria. The plant then absorbs the smaller molecules. (credit a: modification of work by Peter Shanks; credit b: modification of work by Tim Mansfield)

Many swamp plants have adaptations that enable them to thrive in wet areas, where their roots grow submerged underwater. In these aquatic areas, the soil is unstable and little oxygen is available to reach the roots. Trees such as mangroves (*Rhizophora* sp.) growing in coastal waters produce aboveground roots that help support the tree ([link]). Some species of mangroves, as well as cypress trees, have pneumatophores: upward-

growing roots containing pores and pockets of tissue specialized for gas exchange. Wild rice is an aquatic plant with large air spaces in the root cortex. The air-filled tissue—called aerenchyma—provides a path for oxygen to diffuse down to the root tips, which are embedded in oxygen-poor bottom sediments.



The branches of (a) mangrove trees develop aerial roots, which descend to the ground and help to anchor the trees. (b) Cypress trees and some mangrove species have upward-growing roots called pneumatophores that are involved in gas exchange. Aquatic plants such as (c) wild rice have large spaces in the root cortex called aerenchyma, visualized here using scanning electron microscopy. (credit a: modification of work by Roberto Verzo; credit b: modification of work by Duane Burdick; credit c: modification of work by Robert R. Wise)

Note:
Link to Learning

Watch *Venus Flytraps: Jaws of Death*, an extraordinary BBC close-up of the Venus flytrap in action.

https://www.openstaxcollege.org/l/venus_flytrap

Section Summary

Leaves are the main site of photosynthesis. A typical leaf consists of a lamina (the broad part of the leaf, also called the blade) and a petiole (the stalk that attaches the leaf to a stem). The arrangement of leaves on a stem, known as phyllotaxy, enables maximum exposure to sunlight. Each plant species has a characteristic leaf arrangement and form. The pattern of leaf arrangement may be alternate, opposite, or spiral, while leaf form may be simple or compound. Leaf tissue consists of the epidermis, which forms the outermost cell layer, and mesophyll and vascular tissue, which make up the inner portion of the leaf. In some plant species, leaf form is modified to form structures such as tendrils, spines, bud scales, and needles.

Review Questions

Exercise:

Problem: The stalk of a leaf is known as the ________.

a. petiole
b. lamina
c. stipule
d. rachis

Solution:

A

Exercise:

Problem: Leaflets are a characteristic of ________ leaves.

a. alternate
b. whorled
c. compound
d. opposite

Solution:

C

Exercise:

Problem: Cells of the _________ contain chloroplasts.

a. epidermis
b. vascular tissue
c. stomata
d. mesophyll

Solution:

D

Exercise:

Problem:

Which of the following is most likely to be found in a desert environment?

a. broad leaves to capture sunlight
b. spines instead of leaves
c. needle-like leaves
d. wide, flat leaves that can float

Solution:

B

Free Response

Exercise:

Problem:

How do dicots differ from monocots in terms of leaf structure?

Solution:

Monocots have leaves with parallel venation, and dicots have leaves with reticulate, net-like venation.

Exercise:

Problem:

Describe an example of a plant with leaves that are adapted to cold temperatures.

Solution:

Conifers such as spruce, fir, and pine have needle-shaped leaves with sunken stomata, helping to reduce water loss.

Glossary

compound leaf
: leaf in which the leaf blade is subdivided to form leaflets, all attached to the midrib

cuticle
: waxy protective layer on the leaf surface

lamina
: leaf blade

palmately compound leaf
leaf type with leaflets that emerge from a point, resembling the palm of a hand

petiole
stalk of the leaf

phyllotaxy
arrangement of leaves on a stem

pinnately compound leaf
leaf type with a divided leaf blade consisting of leaflets arranged on both sides of the midrib

sessile
leaf without a petiole that is attached directly to the plant stem

simple leaf
leaf type in which the lamina is completely undivided or merely lobed

stipule
small green structure found on either side of the leaf stalk or petiole

venation
pattern of veins in a leaf; may be parallel (as in monocots), reticulate (as in dicots), or dichotomous (as in *Gingko biloba*)

whorled
pattern of leaf arrangement in which three or more leaves are connected at a node

Flowers and Fruit
By the end of this section, you will be able to:

- Describe the main parts of a flower and their purpose

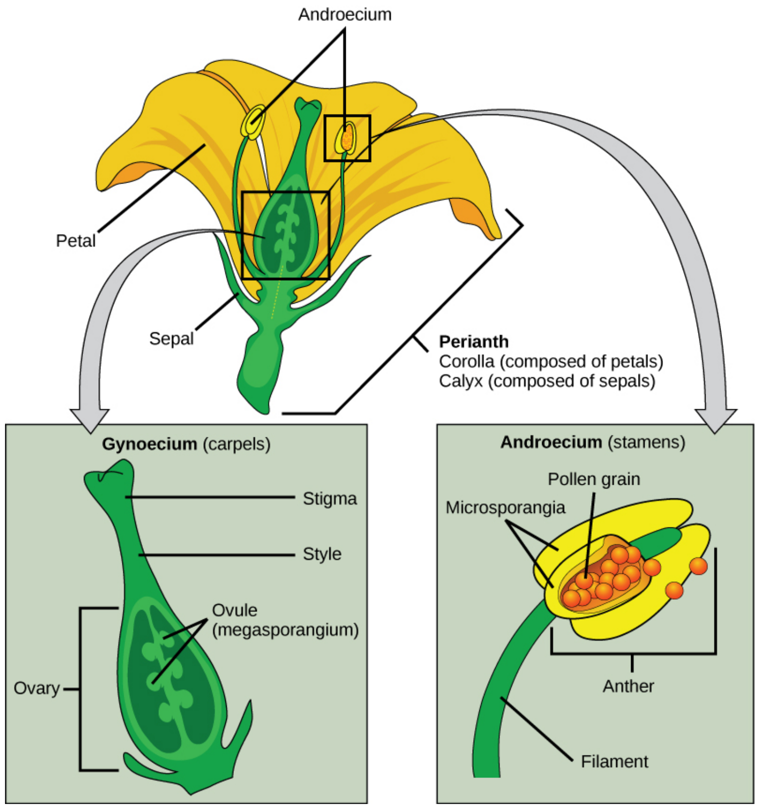
The success of angiosperms is due to two novel reproductive structures: flowers and fruit. The function of the flower is to ensure pollination. Flowers also provide protection for the ovule and developing embryo inside a receptacle. The function of the fruit is seed dispersal. They also protect the developing seed. Different fruit structures or tissues on fruit—such as sweet flesh, wings, parachutes, or spines that grab—reflect the dispersal strategies that help spread seeds.

These flowers grow in a botanical garden border in Bellevue, WA. Flowering plants dominate terrestrial landscapes. The vivid colors of flowers are an adaptation to pollination by animals such as insects and birds. (credit: Myriam Feldman)

Flowers

Flowers are modified leaves, or sporophylls, organized around a central stalk. Although they vary greatly in appearance, all flowers contain the same structures: sepals, petals, carpels, and stamens. The peduncle attaches the flower to the plant. A whorl of **sepals** (collectively called the **calyx**) is located at the base of the peduncle and encloses the unopened floral bud. Sepals are usually photosynthetic organs, although there are some exceptions. For example, the corolla in lilies and tulips consists of three sepals and three petals that look virtually identical. **Petals**, collectively the **corolla**, are located inside the whorl of sepals and often display vivid colors to attract pollinators. Flowers pollinated by wind are usually small, feathery, and visually inconspicuous. Sepals and petals together form the **perianth**. The sexual organs (carpels and stamens) are located at the center of the flower.

As illustrated in [link], styles, stigmas, and ovules constitute the female organ: the **gynoecium** or **carpel**. Flower structure is very diverse, and carpels may be singular, multiple, or fused. Multiple fused carpels comprise a **pistil**. The megaspores and the female gametophytes are produced and protected by the thick tissues of the carpel. A long, thin structure called a **style** leads from the sticky **stigma**, where pollen is deposited, to the **ovary**, enclosed in the carpel. The ovary houses one or more ovules, each of which will develop into a seed upon fertilization. The male reproductive organs, the **stamens** (collectively called the androecium), surround the central carpel. Stamens are composed of a thin stalk called a **filament** and a sac-like structure called the anther. The filament supports the **anther**, where the microspores are produced by meiosis and develop into pollen grains.



This image depicts the structure of a perfect flower. Perfect flowers produce both male and female floral organs. The flower shown has only one carpel, but some flowers have a cluster of carpels. Together, all the carpels make up the gynoecium. (credit: modification of work by Mariana Ruiz Villareal)

Fruit

As the seed develops, the walls of the ovary thicken and form the fruit. The seed forms in an ovary, which also enlarges as the seeds grow. In botany, a fertilized and fully grown, ripened ovary is a fruit. Many foods commonly called vegetables are actually fruit. Eggplants, zucchini, string beans, and bell peppers are all technically fruit because they contain seeds and are

derived from the thick ovary tissue. Acorns are nuts, and winged maple whirligigs (whose botanical name is samara) are also fruit. Botanists classify fruit into more than two dozen different categories, only a few of which are actually fleshy and sweet.

Mature fruit can be fleshy or dry. Fleshy fruit include the familiar berries, peaches, apples, grapes, and tomatoes. Rice, wheat, and nuts are examples of dry fruit. Another distinction is that not all fruits are derived from the ovary. For instance, strawberries are derived from the receptacle and apples from the pericarp, or hypanthium. Some fruits are derived from separate ovaries in a single flower, such as the raspberry. Other fruits, such as the pineapple, form from clusters of flowers. Additionally, some fruits, like watermelon and orange, have rinds. Regardless of how they are formed, fruits are an agent of seed dispersal. The variety of shapes and characteristics reflect the mode of dispersal. Wind carries the light dry fruit of trees and dandelions. Water transports floating coconuts. Some fruits attract herbivores with color or perfume, or as food. Once eaten, tough, undigested seeds are dispersed through the herbivore's feces. Other fruits have burs and hooks to cling to fur and hitch rides on animals.

Section Summary

Angiosperms are the dominant form of plant life in most terrestrial ecosystems, comprising about 90 percent of all plant species. Most crops and ornamental plants are angiosperms. Their success comes from two innovative structures that protect reproduction from variability in the environment: the flower and the fruit. Flowers were derived from modified leaves. The main parts of a flower are the sepals and petals, which protect the reproductive parts: the stamens and the carpels. The stamens produce the male gametes in pollen grains. The carpels contain the female gametes (the eggs inside the ovules), which are within the ovary of a carpel. The walls of the ovary thicken after fertilization, ripening into fruit that ensures dispersal by wind, water, or animals.

Art Connections

Exercise:

Problem:

[link] If a flower lacked a megasporangium, what type of gamete would not form? If the flower lacked a microsporangium, what type of gamete would not form?

Solution:

[link] Without a megasporangium, an egg would not form; without a microsporangium, pollen would not form.

Review Questions

Exercise:

Problem:

Which of the following structures in a flower is not directly involved in reproduction?

a. the style
b. the stamen
c. the sepal
d. the anther

Solution:

C

Exercise:

Problem: Pollen grains develop in which structure?

a. the anther
b. the stigma
c. the filament
d. the carpel

Solution:

A

Exercise:

Problem:

In the course of double fertilization, one sperm cell fuses with the egg and the second one fuses with _________.

a. the synergids
b. the polar nuclei of the center cell
c. the egg as well
d. the antipodal cells

Solution:

B

Exercise:

Problem:

Corn develops from a seedling with a single cotyledon, displays parallel veins on its leaves, and produces monosulcate pollen. It is most likely:

a. a gymnosperm
b. a monocot
c. a eudicot
d. a basal angiosperm

Solution:

B

Free Response

Exercise:

Problem:

Some cycads are considered endangered species and their trade is severely restricted. Customs officials stop suspected smugglers who claim that the plants in their possession are palm trees, not cycads. How would a botanist distinguish between the two types of plants?

Solution:

The resemblance between cycads and palm trees is only superficial. Cycads are gymnosperms and do not bear flowers or fruit. Cycads produce cones: large, female cones that produce naked seeds, and smaller male cones on separate plants. Palms do not.

Exercise:

Problem:

What are the two structures that allow angiosperms to be the dominant form of plant life in most terrestrial ecosystems?

Solution:

Angiosperms are successful because of flowers and fruit. These structures protect reproduction from variability in the environment.

Glossary

anther
: sac-like structure at the tip of the stamen in which pollen grains are produced

Anthophyta
: phylum to which angiosperms belong

basal angiosperms
: a group of plants that probably branched off before the separation of monocots and eudicots

calyx
: whorl of sepals

carpel
: single unit of the pistil

corolla
: collection of petals

cotyledon
: primitive leaf that develop in the zygote; monocots have one cotyledon, and dicots have two cotyledons

dicot
: (also, eudicot) related group of angiosperms whose embryos possess two cotyledons

filament
: thin stalk that links the anther to the base of the flower

gynoecium
: (also, carpel) structure that constitute the female reproductive organ

herbaceous
: grass-like plant noticeable by the absence of woody tissue

monocot
: related group of angiosperms that produce embryos with one cotyledon and pollen with a single ridge

ovary
: chamber that contains and protects the ovule or female megasporangium

perianth

part of the plant consisting of the calyx (sepals) and corolla (petals)

petal
: modified leaf interior to the sepals; colorful petals attract animal pollinators

pistil
: fused group of carpels

sepal
: modified leaf that encloses the bud; outermost structure of a flower

stamen
: structure that contains the male reproductive organs

stigma
: uppermost structure of the carpel where pollen is deposited

style
: long, thin structure that links the stigma to the ovary

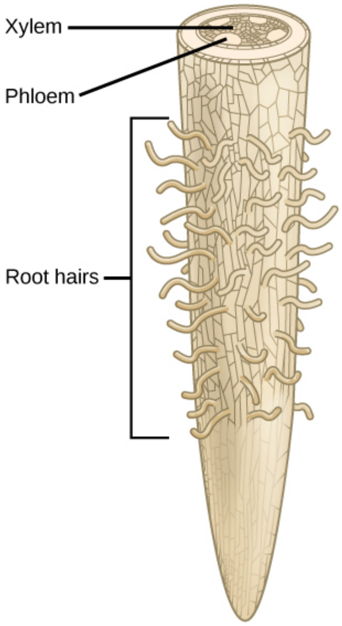
Nutritional Requirements of Plants
By the end of this section, you will be able to:

- Describe how plants obtain nutrients
- List the elements and compounds required for proper plant nutrition
- Describe an essential nutrient

Plants are unique organisms that can absorb nutrients and water through their root system, as well as carbon dioxide from the atmosphere. Soil quality and climate are the major determinants of plant distribution and growth. The combination of soil nutrients, water, and carbon dioxide, along with sunlight, allows plants to grow.

The Chemical Composition of Plants

Since plants require nutrients in the form of elements such as carbon and potassium, it is important to understand the chemical composition of plants. The majority of volume in a plant cell is water; it typically comprises 80 to 90 percent of the plant's total weight. Soil is the water source for land plants, and can be an abundant source of water, even if it appears dry. Plant roots absorb water from the soil through root hairs and transport it up to the leaves through the xylem. As water vapor is lost from the leaves, the process of transpiration and the polarity of water molecules (which enables them to form hydrogen bonds) draws more water from the roots up through the plant to the leaves ([link]). Plants need water to support cell structure, for metabolic functions, to carry nutrients, and for photosynthesis.



Water is absorbed through the root hairs and moves up the xylem to the leaves.

Plant cells need essential substances, collectively called nutrients, to sustain life. Plant nutrients may be composed of either organic or inorganic compounds. An **organic compound** is a chemical compound that contains carbon, such as carbon dioxide obtained from the atmosphere. Carbon that was obtained from atmospheric CO2 composes the majority of the dry mass within most plants. An **inorganic compound** does not contain carbon and is not part of, or produced by, a living organism. Inorganic substances, which form the majority of the soil solution, are commonly called minerals: those required by plants include nitrogen (N) and potassium (K) for structure and regulation.

Essential Nutrients

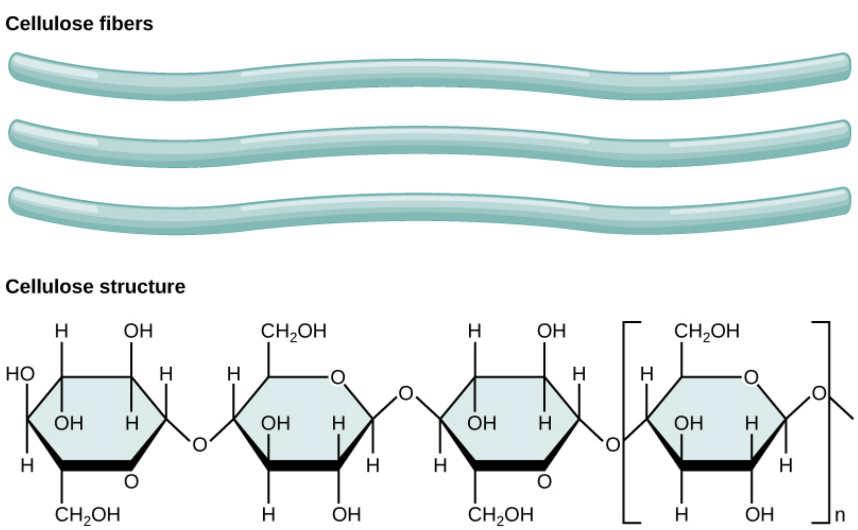
Plants require only light, water and about 20 elements to support all their biochemical needs: these 20 elements are called essential nutrients ([link]). For an element to be regarded as **essential**, three criteria are required: 1) a

plant cannot complete its life cycle without the element; 2) no other element can perform the function of the element; and 3) the element is directly involved in plant nutrition.

Essential Elements for Plant Growth	
Macronutrients	**Micronutrients**
Carbon (C)	Iron (Fe)
Hydrogen (H)	Manganese (Mn)
Oxygen (O)	Boron (B)
Nitrogen (N)	Molybdenum (Mo)
Phosphorus (P)	Copper (Cu)
Potassium (K)	Zinc (Zn)
Calcium (Ca)	Chlorine (Cl)
Magnesium (Mg)	Nickel (Ni)
Sulfur (S)	Cobalt (Co)
	Sodium (Na)
	Silicon (Si)

Macronutrients and Micronutrients

The essential elements can be divided into two groups: macronutrients and micronutrients. Nutrients that plants require in larger amounts are called **macronutrients**. About half of the essential elements are considered macronutrients: carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium, magnesium and sulfur. The first of these macronutrients, carbon (C), is required to form carbohydrates, proteins, nucleic acids, and many other compounds; it is therefore present in all macromolecules. On average, the dry weight (excluding water) of a cell is 50 percent carbon. As shown in [link], carbon is a key part of plant biomolecules.



Cellulose, the main structural component of the plant cell wall, makes up over thirty percent of plant matter. It is the most abundant organic compound on earth.

The next most abundant element in plant cells is nitrogen (N); it is part of proteins and nucleic acids. Nitrogen is also used in the synthesis of some vitamins. Hydrogen and oxygen are macronutrients that are part of many

organic compounds, and also form water. Oxygen is necessary for cellular respiration; plants use oxygen to store energy in the form of ATP. Phosphorus (P), another macromolecule, is necessary to synthesize nucleic acids and phospholipids. As part of ATP, phosphorus enables food energy to be converted into chemical energy through oxidative phosphorylation. Likewise, light energy is converted into chemical energy during photophosphorylation in photosynthesis, and into chemical energy to be extracted during respiration. Sulfur is part of certain amino acids, such as cysteine and methionine, and is present in several coenzymes. Sulfur also plays a role in photosynthesis as part of the electron transport chain, where hydrogen gradients play a key role in the conversion of light energy into ATP. Potassium (K) is important because of its role in regulating stomatal opening and closing. As the openings for gas exchange, stomata help maintain a healthy water balance; a potassium ion pump supports this process.

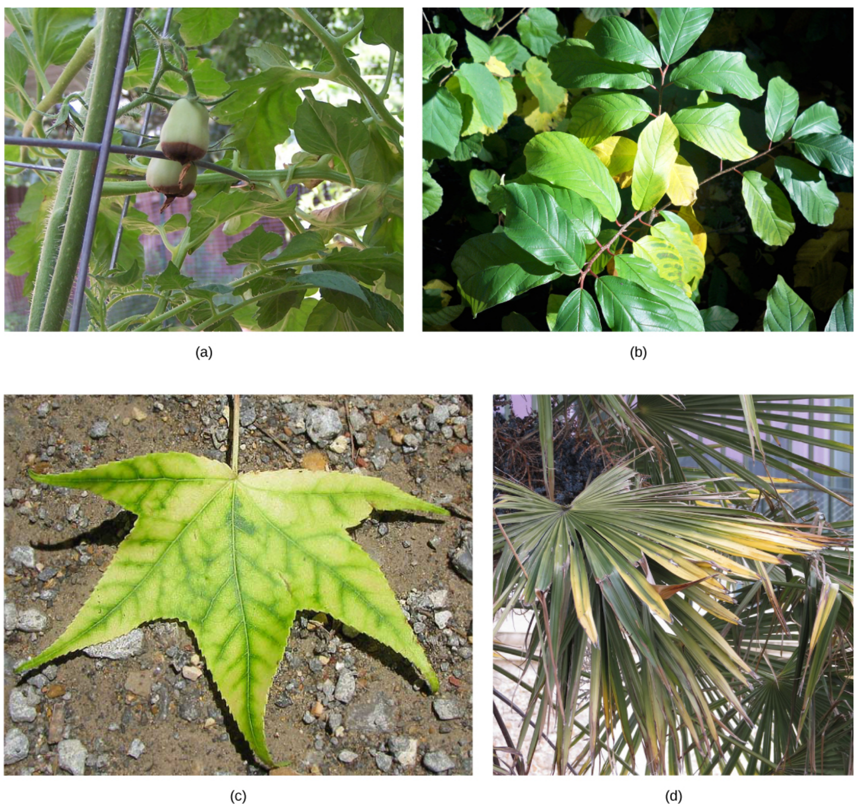
Magnesium (Mg) and calcium (Ca) are also important macronutrients. The role of calcium is twofold: to regulate nutrient transport, and to support many enzyme functions. Magnesium is important to the photosynthetic process. These minerals, along with the micronutrients, which are described below, also contribute to the plant's ionic balance.

In addition to macronutrients, organisms require various elements in small amounts. These **micronutrients**, or trace elements, are present in very small quantities. They include boron (B), chlorine (Cl), manganese (Mn), iron (Fe), zinc (Zn), copper (Cu), molybdenum (Mo), nickel (Ni), silicon (Si), and sodium (Na).

Deficiencies in any of these nutrients—particularly the macronutrients—can adversely affect plant growth ([link]). Depending on the specific nutrient, a lack can cause stunted growth, slow growth, or chlorosis (yellowing of the leaves). Extreme deficiencies may result in leaves showing signs of cell death.

Note:
Link to Learning

Visit this website to participate in an interactive experiment on plant nutrient deficiencies. You can adjust the amounts of N, P, K, Ca, Mg, and Fe that plants receive . . . and see what happens.



Nutrient deficiency is evident in the symptoms these plants show. This (a) grape tomato suffers from blossom end rot caused by calcium deficiency. The yellowing in this (b) *Frangula alnus* results from magnesium deficiency. Inadequate magnesium also leads to (c) intervenal chlorosis, seen here in a sweetgum

leaf. This (d) palm is affected by potassium deficiency. (credit c: modification of work by Jim Conrad; credit d: modification of work by Malcolm Manners)

Note:
Everyday Connection
Hydroponics
Hydroponics is a method of growing plants in a water-nutrient solution instead of soil. Since its advent, hydroponics has developed into a growing process that researchers often use. Scientists who are interested in studying plant nutrient deficiencies can use hydroponics to study the effects of different nutrient combinations under strictly controlled conditions. Hydroponics has also developed as a way to grow flowers, vegetables, and other crops in greenhouse environments. You might find hydroponically grown produce at your local grocery store. Today, many lettuces and tomatoes in your market have been hydroponically grown.

Section Summary

Plants can absorb inorganic nutrients and water through their root system, and carbon dioxide from the environment. The combination of organic compounds, along with water, carbon dioxide, and sunlight, produce the energy that allows plants to grow. Inorganic compounds form the majority of the soil solution. Plants access water though the soil. Water is absorbed by the plant root, transports nutrients throughout the plant, and maintains the structure of the plant. Essential elements are indispensable elements for plant growth. They are divided into macronutrients and micronutrients. The macronutrients plants require are carbon, nitrogen, hydrogen, oxygen, phosphorus, potassium, calcium, magnesium, and sulfur. Important

micronutrients include iron, manganese, boron, molybdenum, copper, zinc, chlorine, nickel, cobalt, silicon and sodium.

Review Questions

Exercise:

Problem:

For an element to be regarded as essential, all of the following criteria must be met, except:

a. No other element can perform the function.
b. The element is directly involved in plant nutrition.
c. The element is inorganic.
d. The plant cannot complete its lifecycle without the element.

Solution:

C

Exercise:

Problem:

The nutrient that is part of carbohydrates, proteins, and nucleic acids, and that forms biomolecules, is _________.

a. nitrogen
b. carbon
c. magnesium
d. iron

Solution:

B

Exercise:

Problem: Most _________ are necessary for enzyme function.

a. micronutrients
b. macronutrients
c. biomolecules
d. essential nutrients

Solution:

A

Exercise:

Problem: What is the main water source for land plants?

a. rain
b. soil
c. biomolecules
d. essential nutrients

Solution:

B

Free Response

Exercise:

Problem:

What type of plant problems result from nitrogen and calcium deficiencies?

Solution:

Deficiencies in these nutrients could result in stunted growth, slow growth, and chlorosis.

Exercise:

Problem:

Research the life of Jan Babtista van Helmont. What did the van Helmont experiment show?

Solution:

van Helmont showed that plants do not consume soil, which is correct. He also thought that plant growth and increased weight resulted from the intake of water, a conclusion that has since been disproven.

Exercise:

Problem:

List two essential macronutrients and two essential nutrients.

Solution:

Answers may vary. Essential macronutrients include carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium, magnesium, and sulfur. Essential micronutrients include iron, manganese, boron, molybdenum, copper, zinc, chlorine, nickel, cobalt, sodium, and silicon.

Glossary

inorganic compound
: chemical compound that does not contain carbon; it is not part of or produced by a living organism

macronutrient
: nutrient that is required in large amounts for plant growth; carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium,

magnesium, and sulfur

micronutrient
: nutrient required in small amounts; also called trace element

organic compound
: chemical compound that contains carbon

Nutritional Adaptations of Plants

By the end of this section, you will be able to:

- Understand the nutritional adaptations of plants
- Describe mycorrhizae
- Explain nitrogen fixation

Plants obtain food in two different ways. Autotrophic plants can make their own food from inorganic raw materials, such as carbon dioxide and water, through photosynthesis in the presence of sunlight. Green plants are included in this group. Some plants, however, are heterotrophic: they are totally parasitic and lacking in chlorophyll. These plants, referred to as holo-parasitic plants, are unable to synthesize organic carbon and draw all of their nutrients from the host plant.

Plants may also enlist the help of microbial partners in nutrient acquisition. Particular species of bacteria and fungi have evolved along with certain plants to create a mutualistic symbiotic relationship with roots. This improves the nutrition of both the plant and the microbe. The formation of nodules in legume plants and mycorrhization can be considered among the nutritional adaptations of plants. However, these are not the only type of adaptations that we may find; many plants have other adaptations that allow them to thrive under specific conditions.

Note:
Link to Learning

This video reviews basic concepts about photosynthesis. In the left panel, click each tab to select a topic for review.

Nitrogen Fixation: Root and Bacteria Interactions

Nitrogen is an important macronutrient because it is part of nucleic acids and proteins. Atmospheric nitrogen, which is the diatomic molecule N_2, or dinitrogen, is the largest pool of nitrogen in terrestrial ecosystems. However, plants cannot take advantage of this nitrogen because they do not have the necessary enzymes to convert it into biologically useful forms. However, nitrogen can be “fixed,” which means that it can be converted to ammonia (NH_3) through biological, physical, or chemical processes. As you have learned, biological nitrogen fixation (BNF) is the conversion of atmospheric nitrogen (N_2) into ammonia (NH_3), exclusively carried out by prokaryotes such as soil bacteria or cyanobacteria. Biological processes contribute 65 percent of the nitrogen used in agriculture. The following equation represents the process:

Equation:

$$N_2 + 16\ \text{ATP}\ +\ 8\ e^-\ +\ 8\ \text{H}^+\ \rightarrow\ 2\text{NH}_3\ +\ 16\ \text{ADP}\ +\ 16\ \text{Pi}\ +\ \text{H}_2$$

The most important source of BNF is the symbiotic interaction between soil bacteria and legume plants, including many crops important to humans ([link]). The NH_3 resulting from fixation can be transported into plant tissue and incorporated into amino acids, which are then made into plant proteins. Some legume seeds, such as soybeans and peanuts, contain high levels of protein, and serve among the most important agricultural sources of protein in the world.

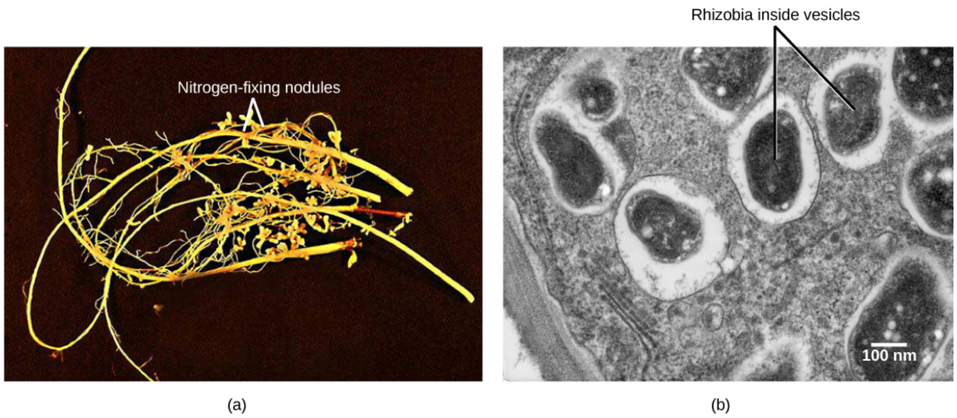
Note:
Art Connection

(a) (b) (c)

Some common edible legumes—like (a) peanuts, (b) beans, and (c) chickpeas—are able to interact symbiotically with soil bacteria that fix nitrogen. (credit a: modification of work by Jules Clancy; credit b: modification of work by USDA)

Farmers often rotate corn (a cereal crop) and soy beans (a legume), planting a field with each crop in alternate seasons. What advantage might this crop rotation confer?

Soil bacteria, collectively called **rhizobia**, symbiotically interact with legume roots to form specialized structures called **nodules**, in which nitrogen fixation takes place. This process entails the reduction of atmospheric nitrogen to ammonia, by means of the enzyme **nitrogenase**. Therefore, using rhizobia is a natural and environmentally friendly way to fertilize plants, as opposed to chemical fertilization that uses a nonrenewable resource, such as natural gas. Through symbiotic nitrogen fixation, the plant benefits from using an endless source of nitrogen from the atmosphere. The process simultaneously contributes to soil fertility because the plant root system leaves behind some of the biologically available nitrogen. As in any symbiosis, both organisms benefit from the interaction: the plant obtains ammonia, and bacteria obtain carbon compounds generated through photosynthesis, as well as a protected niche in which to grow ([link]).



Soybean roots contain (a) nitrogen-fixing nodules. Cells within the nodules are infected with *Bradyrhyzobium japonicum,* a rhizobia or “root-loving” bacterium. The bacteria are encased in (b) vesicles inside the cell, as can be seen in this transmission electron micrograph. (credit a: modification of work by USDA; credit b: modification of work by Louisa Howard, Dartmouth Electron Microscope Facility; scale-bar data from Matt Russell)

Mycorrhizae: The Symbiotic Relationship between Fungi and Roots

A nutrient depletion zone can develop when there is rapid soil solution uptake, low nutrient concentration, low diffusion rate, or low soil moisture. These conditions are very common; therefore, most plants rely on fungi to facilitate the uptake of minerals from the soil. Fungi form symbiotic associations called mycorrhizae with plant roots, in which the fungi actually are integrated into the physical structure of the root. The fungi colonize the living root tissue during active plant growth.

Through mycorrhization, the plant obtains mainly phosphate and other minerals, such as zinc and copper, from the soil. The fungus obtains nutrients, such as sugars, from the plant root ([link]). Mycorrhizae help

increase the surface area of the plant root system because hyphae, which are narrow, can spread beyond the nutrient depletion zone. Hyphae can grow into small soil pores that allow access to phosphorus that would otherwise be unavailable to the plant. The beneficial effect on the plant is best observed in poor soils. The benefit to fungi is that they can obtain up to 20 percent of the total carbon accessed by plants. Mycorrhizae functions as a physical barrier to pathogens. It also provides an induction of generalized host defense mechanisms, and sometimes involves production of antibiotic compounds by the fungi.

Root tips proliferate in the presence of mycorrhizal infection, which appears as off-white fuzz in this image. (credit: modification of work by Nilsson et al., BMC Bioinformatics 2005)

There are two types of mycorrhizae: ectomycorrhizae and endomycorrhizae. Ectomycorrhizae form an extensive dense sheath around the roots, called a mantle. Hyphae from the fungi extend from the mantle into the soil, which increases the surface area for water and mineral absorption. This type of mycorrhizae is found in forest trees, especially

conifers, birches, and oaks. Endomycorrhizae, also called arbuscular mycorrhizae, do not form a dense sheath over the root. Instead, the fungal mycelium is embedded within the root tissue. Endomycorrhizae are found in the roots of more than 80 percent of terrestrial plants.

Nutrients from Other Sources

Some plants cannot produce their own food and must obtain their nutrition from outside sources. This may occur with plants that are parasitic or saprophytic. Some plants are mutualistic symbionts, epiphytes, or insectivorous.

Plant Parasites

A **parasitic plant** depends on its host for survival. Some parasitic plants have no leaves. An example of this is the dodder ([link]), which has a weak, cylindrical stem that coils around the host and forms suckers. From these suckers, cells invade the host stem and grow to connect with the vascular bundles of the host. The parasitic plant obtains water and nutrients through these connections. The plant is a total parasite (a holoparasite) because it is completely dependent on its host. Other parasitic plants (hemiparasites) are fully photosynthetic and only use the host for water and minerals. There are about 4,100 species of parasitic plants.

The dodder is a holoparasite that penetrates the host’s vascular tissue and diverts nutrients for its own growth. Note that the vines of the dodder, which has white flowers, are beige. The dodder has no chlorophyll and cannot produce its own food. (credit: "Lalithamba"/Flickr)

Saprophytes

A **saprophyte** is a plant that does not have chlorophyll and gets its food from dead matter, similar to bacteria and fungi (note that fungi are often called saprophytes, which is incorrect, because fungi are not plants). Plants like these use enzymes to convert organic food materials into simpler forms from which they can absorb nutrients ([link]). Most saprophytes do not directly digest dead matter: instead, they parasitize fungi that digest dead matter, or are mycorrhizal, ultimately obtaining photosynthate from a fungus that derived photosynthate from its host. Saprophytic plants are uncommon; only a few species are described.

Saprophytes, like this Dutchmen’s pipe (*Monotropa hypopitys)*, obtain their food from dead matter and do not have chlorophyll. (credit: modification of work by Iwona Erskine-Kellie)

Symbionts

A **symbiont** is a plant in a symbiotic relationship, with special adaptations such as mycorrhizae or nodule formation. Fungi also form symbiotic associations with cyanobacteria and green algae (called lichens). Lichens can sometimes be seen as colorful growths on the surface of rocks and trees ([link]). The algal partner (phycobiont) makes food autotrophically, some of which it shares with the fungus; the fungal partner (mycobiont) absorbs water and minerals from the environment, which are made available to the green alga. If one partner was separated from the other, they would both die.

Lichens, which often have symbiotic relationships with other plants, can sometimes be found growing on trees. (credit: "benketaro"/Flickr)

Epiphytes

An **epiphyte** is a plant that grows on other plants, but is not dependent upon the other plant for nutrition ([link]). Epiphytes have two types of roots: clinging aerial roots, which absorb nutrients from humus that accumulates in the crevices of trees; and aerial roots, which absorb moisture from the atmosphere.

These epiphyte plants grow in the main greenhouse of the *Jardin des Plantes* in Paris.

Insectivorous Plants

An **insectivorous** plant has specialized leaves to attract and digest insects. The Venus flytrap is popularly known for its insectivorous mode of nutrition, and has leaves that work as traps ([link]). The minerals it obtains from prey compensate for those lacking in the boggy (low pH) soil of its native North Carolina coastal plains. There are three sensitive hairs in the center of each half of each leaf. The edges of each leaf are covered with long spines. Nectar secreted by the plant attracts flies to the leaf. When a fly touches the sensory hairs, the leaf immediately closes. Next, fluids and enzymes break down the prey and minerals are absorbed by the leaf. Since this plant is popular in the horticultural trade, it is threatened in its original habitat.

A Venus flytrap has specialized leaves to trap insects. (credit: "Selena N. B. H."/Flickr)

Section Summary

Atmospheric nitrogen is the largest pool of available nitrogen in terrestrial ecosystems. However, plants cannot use this nitrogen because they do not have the necessary enzymes. Biological nitrogen fixation (BNF) is the conversion of atmospheric nitrogen to ammonia. The most important source of BNF is the symbiotic interaction between soil bacteria and legumes. The bacteria form nodules on the legume’s roots in which nitrogen fixation takes place. Fungi form symbiotic associations (mycorrhizae) with plants, becoming integrated into the physical structure of the root. Through mycorrhization, the plant obtains minerals from the soil and the fungus obtains photosynthate from the plant root. Ectomycorrhizae form an extensive dense sheath around the root, while endomycorrhizae are embedded within the root tissue. Some plants—parasites, saprophytes, symbionts, epiphytes, and insectivores—have evolved adaptations to obtain their organic or mineral nutrition from various sources.

Art Connections

Exercise:

Problem:

[link] Farmers often rotate corn (a cereal crop) and soy beans (a legume) planting a field with each crop in alternate seasons. What advantage might this crop rotation confer?

Solution:

[link] Soybeans are able to fix nitrogen in their roots, which are not harvested at the end of the growing season. The belowground nitrogen can be used in the next season by the corn.

Review Questions

Exercise:

Problem:

Which process produces an inorganic compound that plants can easily use?

a. photosynthesis
b. nitrogen fixation
c. mycorrhization
d. Calvin cycle

Solution:

B

Exercise:

Problem:

Through mycorrhization, a plant obtains important nutrients such as _________.

a. phosphorus, zinc, and copper
b. phosphorus, zinc, and calcium
c. nickel, calcium, and zinc
d. all of the above

Solution:

A

Exercise:

Problem:

What term describes a plant that requires nutrition from a living host plant?

a. parasite
b. saprophyte
c. epiphyte
d. insectivorous

Solution:

A

Exercise:

Problem:

What is the term for the symbiotic association between fungi and cyanobacteria?

a. lichen
b. mycorrhizae
c. epiphyte
d. nitrogen-fixing nodule

Solution:

A

Free Response

Exercise:

Problem:

Why is biological nitrogen fixation an environmentally friendly way of fertilizing plants?

Solution:

Because it is natural and does not require use of a nonrenewable resource, such as natural gas.

Exercise:

Problem:

What is the main difference, from an energy point of view, between photosynthesis and biological nitrogen fixation?

Solution:

Photosynthesis harvests and stores energy, whereas biological nitrogen fixation requires energy.

Exercise:

Problem: Why is a root nodule a nutritional adaptation of a plant?

Solution:

A nodule results from the symbiosis between a plant and bacterium. Within nodules, the process of nitrogen fixation allows the plant to obtain nitrogen from the air.

Glossary

epiphyte
: plant that grows on other plants but is not dependent upon other plants for nutrition

insectivorous plant
: plant that has specialized leaves to attract and digest insects

nitrogenase
: enzyme that is responsible for the reduction of atmospheric nitrogen to ammonia

nodules
: specialized structures that contain *Rhizobia* bacteria where nitrogen fixation takes place

parasitic plant
: plant that is dependent on its host for survival

rhizobia
: soil bacteria that symbiotically interact with legume roots to form nodules and fix nitrogen

saprophyte
: plant that does not have chlorophyll and gets its food from dead matter

symbiont
: plant in a symbiotic relationship with bacteria or fungi

Transport of Water and Solutes in Plants
By the end of this section, you will be able to:

- Define water potential and explain how it is influenced by solutes, pressure, gravity, and the matric potential
- Describe how water potential, evapotranspiration, and stomatal regulation influence how water is transported in plants
- Explain how photosynthates are transported in plants

The structure of plant roots, stems, and leaves facilitates the transport of water, nutrients, and photosynthates throughout the plant. The phloem and xylem are the main tissues responsible for this movement. Water potential, evapotranspiration, and stomatal regulation influence how water and nutrients are transported in plants. To understand how these processes work, we must first understand the energetics of water potential.

Water Potential

Plants are phenomenal hydraulic engineers. Using only the basic laws of physics and the simple manipulation of potential energy, plants can move water to the top of a 116-meter-tall tree ([link]**a**). Plants can also use hydraulics to generate enough force to split rocks and buckle sidewalks ([link]**b**). Plants achieve this because of water potential.

With heights nearing 116 meters, (a) coastal redwoods (*Sequoia sempervirens*) are the tallest trees in the world. Plant roots can easily generate enough force to (b) buckle and break concrete sidewalks, much to the dismay of homeowners and city maintenance departments. (credit a: modification of work by Bernt Rostad; credit b: modification of work by Pedestrians Educating Drivers on Safety, Inc.)

Water potential is a measure of the potential energy in water. Plant physiologists are not interested in the energy in any one particular aqueous system, but are very interested in water movement between two systems. In practical terms, therefore, water potential is the difference in potential energy between a given water sample and pure water (at atmospheric pressure and ambient temperature). Water potential is denoted by the Greek letter ψ (*psi*) and is expressed in units of pressure (pressure is a form of energy) called **megapascals** (MPa). The potential of pure water ($\Psi_w^{\text{pure H2O}}$) is, by convenience of definition, designated a value of zero (even though pure water contains plenty of potential energy, that energy is ignored). Water potential values for the water in a plant root, stem, or leaf are therefore expressed relative to $\Psi_w^{\text{pure H2O}}$.

The water potential in plant solutions is influenced by solute concentration, pressure, gravity, and factors called matrix effects. Water potential can be broken down into its individual components using the following equation:

Equation:

$$\Psi_{\text{system}} = \Psi_{\text{total}} = \Psi_{\text{s}} + \Psi_{\text{p}} + \Psi_{\text{g}} + \Psi_{\text{m}}$$

where Ψ_s, Ψ_p, Ψ_g, and Ψ_m refer to the solute, pressure, gravity, and matric potentials, respectively. “System” can refer to the water potential of the soil water (Ψ^{soil}), root water (Ψ^{root}), stem water (Ψ^{stem}), leaf water (Ψ^{leaf}) or the water in the atmosphere ($\Psi^{\text{atmosphere}}$): whichever aqueous system is under consideration. As the individual components change, they raise or lower the

total water potential of a system. When this happens, water moves to equilibrate, moving from the system or compartment with a higher water potential to the system or compartment with a lower water potential. This brings the difference in water potential between the two systems ($\Delta\Psi$) back to zero ($\Delta\Psi = 0$). Therefore, for water to move through the plant from the soil to the air (a process called transpiration), Ψ^{soil} must be $> \Psi^{root} > \Psi^{stem} > \Psi^{leaf} > \Psi^{atmosphere}$.

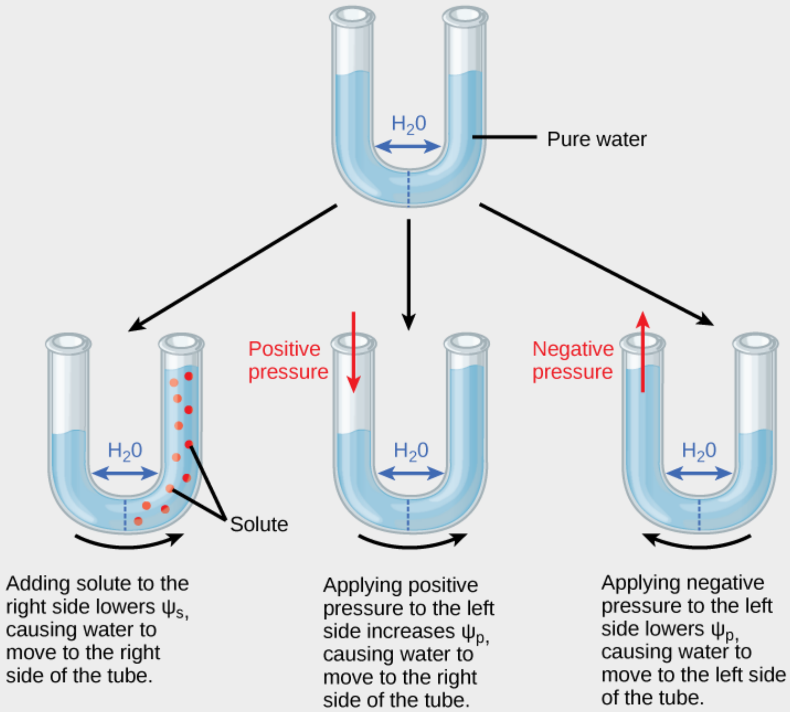
Water only moves in response to $\Delta\Psi$, not in response to the individual components. However, because the individual components influence the total Ψ_{system}, by manipulating the individual components (especially Ψ_s), a plant can control water movement.

Solute Potential

Solute potential (Ψ_s), also called osmotic potential, is negative in a plant cell and zero in distilled water. Typical values for cell cytoplasm are –0.5 to –1.0 MPa. Solutes reduce water potential (resulting in a negative Ψ_w) by consuming some of the potential energy available in the water. Solute molecules can dissolve in water because water molecules can bind to them via hydrogen bonds; a hydrophobic molecule like oil, which cannot bind to water, cannot go into solution. The energy in the hydrogen bonds between solute molecules and water is no longer available to do work in the system because it is tied up in the bond. In other words, the amount of available potential energy is reduced when solutes are added to an aqueous system. Thus, Ψ_s decreases with increasing solute concentration. Because Ψ_s is one of the four components of Ψ_{system} or Ψ_{total}, a decrease in Ψ_s will cause a decrease in Ψ_{total}. The internal water potential of a plant cell is more negative than pure water because of the cytoplasm’s high solute content ([link]). Because of this difference in water potential water will move from the soil into a plant’s root cells via the process of osmosis. This is why solute potential is sometimes called osmotic potential.

Plant cells can metabolically manipulate Ψ_s (and by extension, Ψ_{total}) by adding or removing solute molecules. Therefore, plants have control over Ψ_{total} via their ability to exert metabolic control over Ψ_s.

Note:



In this example with a semipermeable membrane between two aqueous systems, water will move from a region of higher to lower water potential until equilibrium is reached. Solutes (Ψ_s), pressure (Ψ_p), and gravity (Ψ_g) influence total water potential for each side of the tube ($\Psi_{total}{}^{\text{right or left}}$), and therefore, the difference between Ψ_{total} on each side ($\Delta\Psi$). (Ψ_m , the potential due to interaction of water with solid substrates, is ignored in this example because glass is not especially hydrophilic). Water moves in response to the difference in water potential between two systems (the left and right sides of the tube).

Positive water potential is placed on the left side of the tube by increasing Ψ_p such that the water level rises on the right side. Could you equalize the

water level on each side of the tube by adding solute, and if so, how?

Pressure Potential

Pressure potential (Ψ_p), also called turgor potential, may be positive or negative ([link]). Because pressure is an expression of energy, the higher the pressure, the more potential energy in a system, and vice versa. Therefore, a positive Ψp (compression) increases Ψ_{total}, and a negative Ψ_p (tension) decreases Ψ_{total}. Positive pressure inside cells is contained by the cell wall, producing turgor pressure. Pressure potentials are typically around 0.6–0.8 MPa, but can reach as high as 1.5 MPa in a well-watered plant. A Ψ_p of 1.5 MPa equates to 210 pounds per square inch (1.5 MPa x 140 lb in^{-2} MPa^{-1} = 210 lb/in^{-2}). As a comparison, most automobile tires are kept at a pressure of 30–34 psi. An example of the effect of turgor pressure is the wilting of leaves and their restoration after the plant has been watered ([link]). Water is lost from the leaves via transpiration (approaching Ψ_p = 0 MPa at the wilting point) and restored by uptake via the roots.

A plant can manipulate Ψ_p via its ability to manipulate Ψ_s and by the process of osmosis. If a plant cell increases the cytoplasmic solute concentration, Ψ_s will decline, Ψ_{total} will decline, the ΔΨ between the cell and the surrounding tissue will decline, water will move into the cell by osmosis, and Ψ_p will increase. Ψ_p is also under indirect plant control via the opening and closing of stomata. Stomatal openings allow water to evaporate from the leaf, reducing Ψ_p and Ψ_{total} of the leaf and increasing ii between the water in the leaf and the petiole, thereby allowing water to flow from the petiole into the leaf.

When (a) total water potential (Ψ_{total}) is lower outside the cells than inside, water moves out of the cells and the plant wilts. When (b) the total water potential is higher outside the plant cells than inside, water moves into the cells, resulting in turgor pressure (Ψ_p) and keeping the plant erect. (credit: modification of work by Victor M. Vicente Selvas)

Gravity Potential

Gravity potential (Ψ_g) is always negative to zero in a plant with no height. It always removes or consumes potential energy from the system. The force of gravity pulls water downwards to the soil, reducing the total amount of potential energy in the water in the plant (Ψ_{total}). The taller the plant, the taller the water column, and the more influential Ψ_g becomes. On a cellular scale and in short plants, this effect is negligible and easily ignored. However, over the height of a tall tree like a giant coastal redwood, the gravitational pull of -0.1 MPa m^{-1} is equivalent to an extra 1 MPa of resistance that must be overcome for water to reach the leaves of the tallest trees. Plants are unable to manipulate Ψ_g.

Matric Potential

Matric potential (Ψ_m) is always negative to zero. In a dry system, it can be as low as –2 MPa in a dry seed, and it is zero in a water-saturated system. The binding of water to a matrix always removes or consumes potential energy from the system. Ψ_m is similar to solute potential because it involves tying up the energy in an aqueous system by forming hydrogen bonds between the water and some other component. However, in solute potential, the other components are soluble, hydrophilic solute molecules, whereas in Ψ_m, the other components are insoluble, hydrophilic molecules of the plant cell wall. Every plant cell has a cellulosic cell wall and the cellulose in the cell walls is hydrophilic, producing a matrix for adhesion of water: hence the name matric potential. Ψ_m is very large (negative) in dry tissues such as seeds or drought-affected soils. However, it quickly goes to zero as the seed takes up water or the soil hydrates. Ψ_m cannot be manipulated by the plant and is typically ignored in well-watered roots, stems, and leaves.

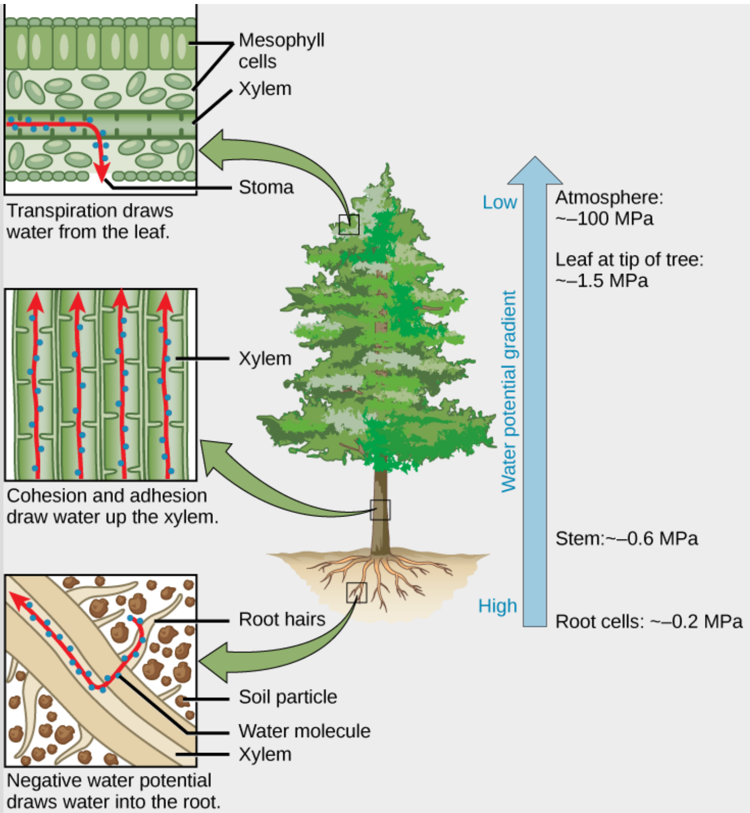
Movement of Water and Minerals in the Xylem

Solutes, pressure, gravity, and matric potential are all important for the transport of water in plants. Water moves from an area of higher total water potential (higher Gibbs free energy) to an area of lower total water potential. Gibbs free energy is the energy associated with a chemical reaction that can be used to do work. This is expressed as $\Delta\Psi$.

Transpiration is the loss of water from the plant through evaporation at the leaf surface. It is the main driver of water movement in the xylem. Transpiration is caused by the evaporation of water at the leaf–atmosphere interface; it creates negative pressure (tension) equivalent to –2 MPa at the leaf surface. This value varies greatly depending on the vapor pressure deficit, which can be negligible at high relative humidity (RH) and substantial at low RH. Water from the roots is pulled up by this tension. At night, when stomata shut and transpiration stops, the water is held in the stem and leaf by the adhesion of water to the cell walls of the xylem vessels and tracheids, and the cohesion of water molecules to each other. This is called the cohesion–tension theory of sap ascent.

Inside the leaf at the cellular level, water on the surface of mesophyll cells saturates the cellulose microfibrils of the primary cell wall. The leaf contains many large intercellular air spaces for the exchange of oxygen for carbon dioxide, which is required for photosynthesis. The wet cell wall is exposed to this leaf internal air space, and the water on the surface of the cells evaporates into the air spaces, decreasing the thin film on the surface of the mesophyll cells. This decrease creates a greater tension on the water in the mesophyll cells ([link]), thereby increasing the pull on the water in the xylem vessels. The xylem vessels and tracheids are structurally adapted to cope with large changes in pressure. Rings in the vessels maintain their tubular shape, much like the rings on a vacuum cleaner hose keep the hose open while it is under pressure. Small perforations between vessel elements reduce the number and size of gas bubbles that can form via a process called cavitation. The formation of gas bubbles in xylem interrupts the continuous stream of water from the base to the top of the plant, causing a break termed an embolism in the flow of xylem sap. The taller the tree, the greater the tension forces needed to pull water, and the more cavitation events. In larger trees, the resulting embolisms can plug xylem vessels, making them non-functional.

Note:
Art Connection



The cohesion–tension theory of sap ascent is shown. Evaporation from the mesophyll cells produces a negative water potential gradient that causes water to move upwards from the roots through the xylem.

Which of the following statements is false?

a. Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the xylem. Transpiration draws water from the leaf.
b. Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the phloem. Transpiration draws water from the leaf.
c. Water potential decreases from the roots to the top of the plant.
d. Water enters the plants through root hairs and exits through stoma.

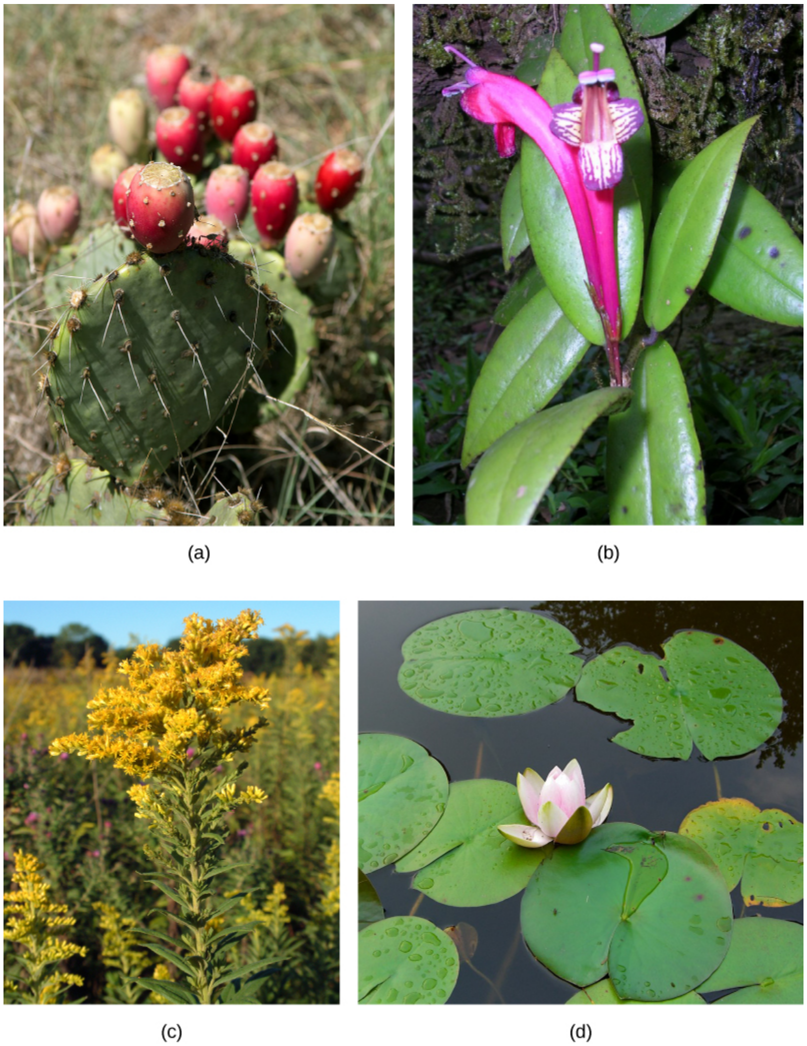
Transpiration—the loss of water vapor to the atmosphere through stomata—is a passive process, meaning that metabolic energy in the form of ATP is not required for water movement. The energy driving transpiration is the difference in energy between the water in the soil and the water in the atmosphere. However, transpiration is tightly controlled.

Control of Transpiration

The atmosphere to which the leaf is exposed drives transpiration, but also causes massive water loss from the plant. Up to 90 percent of the water taken up by roots may be lost through transpiration.

Leaves are covered by a waxy **cuticle** on the outer surface that prevents the loss of water. Regulation of transpiration, therefore, is achieved primarily through the opening and closing of stomata on the leaf surface. Stomata are surrounded by two specialized cells called guard cells, which open and close in response to environmental cues such as light intensity and quality, leaf water status, and carbon dioxide concentrations. Stomata must open to allow air containing carbon dioxide and oxygen to diffuse into the leaf for photosynthesis and respiration. When stomata are open, however, water vapor is lost to the external environment, increasing the rate of transpiration. Therefore, plants must maintain a balance between efficient photosynthesis and water loss.

Plants have evolved over time to adapt to their local environment and reduce transpiration([link]). Desert plant (xerophytes) and plants that grow on other plants (epiphytes) have limited access to water. Such plants usually have a much thicker waxy cuticle than those growing in more moderate, well-watered environments (mesophytes). Aquatic plants (hydrophytes) also have their own set of anatomical and morphological leaf adaptations.



Plants are suited to their local environment. (a) Xerophytes, like this prickly pear cactus (*Opuntia sp.*) and (b) epiphytes such as this tropical *Aeschynanthus perrottetii* have adapted to very limited water resources. The leaves of a prickly pear are modified into spines, which lowers the surface-to-volume ratio and reduces water loss. Photosynthesis takes place in the stem, which also stores water. (b) *A. perottetii* leaves have a waxy cuticle that prevents water loss. (c) Goldenrod (*Solidago sp.*) is a mesophyte, well suited for moderate environments. (d)

Hydrophytes, like this fragrant water lily (*Nymphaea odorata*), are adapted to thrive in aquatic environments. (credit a: modification of work by Jon Sullivan; credit b: modification of work by L. Shyamal/Wikimedia Commons; credit c: modification of work by Huw Williams; credit d: modification of work by Jason Hollinger)

Xerophytes and epiphytes often have a thick covering of trichomes or of stomata that are sunken below the leaf's surface. Trichomes are specialized hair-like epidermal cells that secrete oils and substances. These adaptations impede air flow across the stomatal pore and reduce transpiration. Multiple epidermal layers are also commonly found in these types of plants.

Transportation of Photosynthates in the Phloem

Plants need an energy source to grow. In seeds and bulbs, food is stored in polymers (such as starch) that are converted by metabolic processes into sucrose for newly developing plants. Once green shoots and leaves are growing, plants are able to produce their own food by photosynthesizing. The products of photosynthesis are called photosynthates, which are usually in the form of simple sugars such as sucrose.

Structures that produce photosynthates for the growing plant are referred to as **sources**. Sugars produced in sources, such as leaves, need to be delivered to growing parts of the plant via the phloem in a process called **translocation**. The points of sugar delivery, such as roots, young shoots, and developing seeds, are called **sinks**. Seeds, tubers, and bulbs can be either a source or a sink, depending on the plant's stage of development and the season.

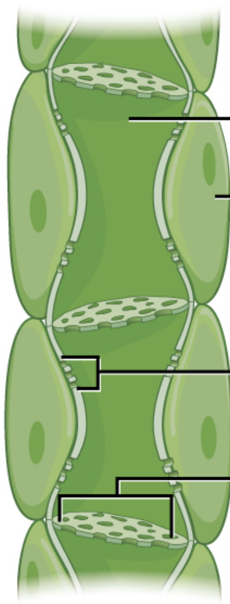
The products from the source are usually translocated to the nearest sink through the phloem. For example, the highest leaves will send

photosynthates upward to the growing shoot tip, whereas lower leaves will direct photosynthates downward to the roots. Intermediate leaves will send products in both directions, unlike the flow in the xylem, which is always unidirectional (soil to leaf to atmosphere). The pattern of photosynthate flow changes as the plant grows and develops. Photosynthates are directed primarily to the roots early on, to shoots and leaves during vegetative growth, and to seeds and fruits during reproductive development. They are also directed to tubers for storage.

Translocation: Transport from Source to Sink

Photosynthates, such as sucrose, are produced in the mesophyll cells of photosynthesizing leaves. From there they are translocated through the phloem to where they are used or stored. Mesophyll cells are connected by cytoplasmic channels called plasmodesmata. Photosynthates move through these channels to reach phloem sieve-tube elements (STEs) in the vascular bundles. From the mesophyll cells, the photosynthates are loaded into the phloem STEs. The sucrose is actively transported against its concentration gradient (a process requiring ATP) into the phloem cells using the electrochemical potential of the proton gradient. This is coupled to the uptake of sucrose with a carrier protein called the sucrose-H^+ symporter.

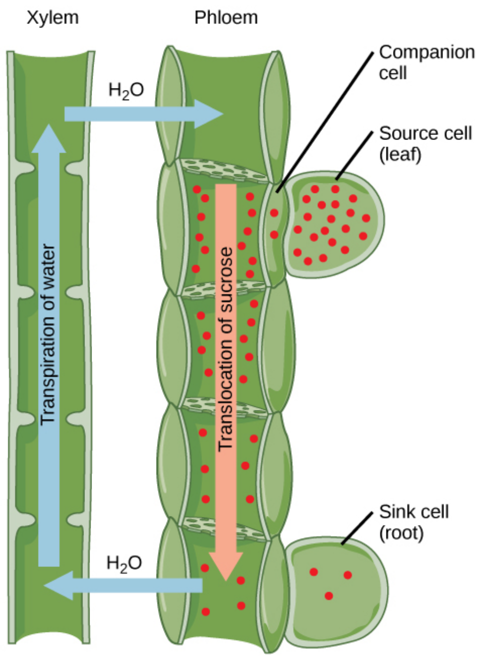
Phloem STEs have reduced cytoplasmic contents, and are connected by a sieve plate with pores that allow for pressure-driven bulk flow, or translocation, of phloem sap. Companion cells are associated with STEs. They assist with metabolic activities and produce energy for the STEs ([link]).



Phloem is comprised of cells called sieve-tube elements. Phloem sap travels through perforations called sieve tube plates. Neighboring companion cells carry out metabolic functions for the sieve-tube elements and provide them with energy. Lateral sieve areas connect the sieve-tube elements to the companion cells.

Once in the phloem, the photosynthates are translocated to the closest sink. Phloem sap is an aqueous solution that contains up to 30 percent sugar, minerals, amino acids, and plant growth regulators. The high percentage of sugar decreases Ψ_s, which decreases the total water potential and causes

water to move by osmosis from the adjacent xylem into the phloem tubes, thereby increasing pressure. This increase in total water potential causes the bulk flow of phloem from source to sink ([link]). Sucrose concentration in the sink cells is lower than in the phloem STEs because the sink sucrose has been metabolized for growth, or converted to starch for storage or other polymers, such as cellulose, for structural integrity. Unloading at the sink end of the phloem tube occurs by either diffusion or active transport of sucrose molecules from an area of high concentration to one of low concentration. Water diffuses from the phloem by osmosis and is then transpired or recycled via the xylem back into the phloem sap.



Sucrose is actively transported from source cells into companion cells and then into the sieve-tube elements. This reduces the water potential, which causes water to enter the phloem from the xylem.

The resulting positive pressure forces the sucrose-water mixture down toward the roots, where sucrose is unloaded. Transpiration causes water to return to the leaves through the xylem vessels.

Section Summary

Water potential (Ψ) is a measure of the difference in potential energy between a water sample and pure water. The water potential in plant solutions is influenced by solute concentration, pressure, gravity, and matric potential. Water potential and transpiration influence how water is transported through the xylem in plants. These processes are regulated by stomatal opening and closing. Photosynthates (mainly sucrose) move from sources to sinks through the plant's phloem. Sucrose is actively loaded into the sieve-tube elements of the phloem. The increased solute concentration causes water to move by osmosis from the xylem into the phloem. The positive pressure that is produced pushes water and solutes down the pressure gradient. The sucrose is unloaded into the sink, and the water returns to the xylem vessels.

Art Connections

Exercise:

Problem:

[link] Positive water potential is placed on the left side of the tube by increasing Ψ_p such that the water level rises on the right side. Could you equalize the water level on each side of the tube by adding solute, and if so, how?

Solution:

[link] Yes, you can equalize the water level by adding the solute to the left side of the tube such that water moves toward the left until the water levels are equal.

Exercise:

Problem: [link] Which of the following statements is false?

a. Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the xylem. Transpiration draws water from the leaf.
b. Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the phloem. Transpiration draws water from the leaf.
c. Water potential decreases from the roots to the top of the plant.
d. Water enters the plants through root hairs and exits through stoma.

Solution:

[link] B.

Review Questions

Exercise:

Problem: When stomata open, what occurs?

a. Water vapor is lost to the external environment, increasing the rate of transpiration.
b. Water vapor is lost to the external environment, decreasing the rate of transpiration.

c. Water vapor enters the spaces in the mesophyll, increasing the rate of transpiration.
d. Water vapor enters the spaces in the mesophyll, increasing the rate of transpiration.

Solution:

A

Exercise:

Problem:

Which cells are responsible for the movement of photosynthates through a plant?

a. tracheids, vessel elements
b. tracheids, companion cells
c. vessel elements, companion cells
d. sieve-tube elements, companion cells

Solution:

D

Free Response

Exercise:

Problem:

The process of bulk flow transports fluids in a plant. Describe the two main bulk flow processes.

Solution:

The process of bulk flow moves water up the xylem and moves photosynthates (solutes) up and down the phloem.

Glossary

cuticle
: waxy covering on the outside of the leaf and stem that prevents the loss of water

megapascal (MPa)
: pressure units that measure water potential

sink
: growing parts of a plant, such as roots and young leaves, which require photosynthate

source
: organ that produces photosynthate for a plant

translocation
: mass transport of photosynthates from source to sink in vascular plants

transpiration
: loss of water vapor to the atmosphere through stomata

water potential (Ψ_w)
: the potential energy of a water solution per unit volume in relation to pure water at atmospheric pressure and ambient temperature

Plant Sensory Systems and Responses
By the end of this section, you will be able to:

- Describe how red and blue light affect plant growth and metabolic activities
- Discuss gravitropism
- Understand how hormones affect plant growth and development
- Describe thigmotropism, thigmonastism, and thigmogenesis
- Explain how plants defend themselves from predators and respond to wounds

Animals can respond to environmental factors by moving to a new location. Plants, however, are rooted in place and must respond to the surrounding environmental factors. Plants have sophisticated systems to detect and respond to light, gravity, temperature, and physical touch. Receptors sense environmental factors and relay the information to effector systems—often through intermediate chemical messengers—to bring about plant responses.

Plant Responses to Light

Plants have a number of sophisticated uses for light that go far beyond their ability to photosynthesize low-molecular-weight sugars using only carbon dioxide, light, and water. **Photomorphogenesis** is the growth and development of plants in response to light. It allows plants to optimize their use of light and space. **Photoperiodism** is the ability to use light to track time. Plants can tell the time of day and time of year by sensing and using various wavelengths of sunlight. **Phototropism** is a directional response that allows plants to grow towards, or even away from, light.

The sensing of light in the environment is important to plants; it can be crucial for competition and survival. The response of plants to light is mediated by different photoreceptors, which are comprised of a protein covalently bonded to a light-absorbing pigment called a **chromophore**. Together, the two are called a chromoprotein.

The red/far-red and violet-blue regions of the visible light spectrum trigger structural development in plants. Sensory photoreceptors absorb light in

these particular regions of the visible light spectrum because of the quality of light available in the daylight spectrum. In terrestrial habitats, light absorption by chlorophylls peaks in the blue and red regions of the spectrum. As light filters through the canopy and the blue and red wavelengths are absorbed, the spectrum shifts to the far-red end, shifting the plant community to those plants better adapted to respond to far-red light. Blue-light receptors allow plants to gauge the direction and abundance of sunlight, which is rich in blue–green emissions. Water absorbs red light, which makes the detection of blue light essential for algae and aquatic plants.

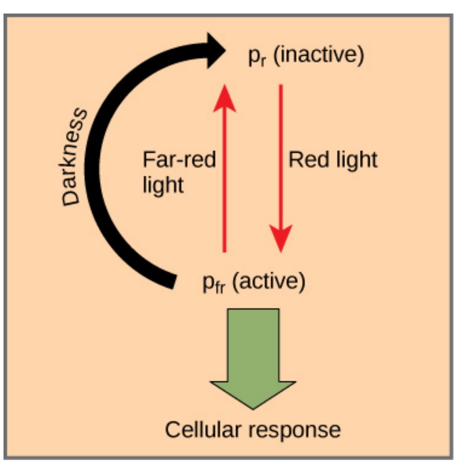
The Phytochrome System and the Red/Far-Red Response

The **phytochromes** are a family of chromoproteins with a linear tetrapyrrole chromophore, similar to the ringed tetrapyrrole light-absorbing head group of chlorophyll. Phytochromes have two photo-interconvertible forms: Pr and Pfr. Pr absorbs red light (~667 nm) and is immediately converted to Pfr. Pfr absorbs far-red light (~730 nm) and is quickly converted back to Pr. Absorption of red or far-red light causes a massive change to the shape of the chromophore, altering the conformation and activity of the phytochrome protein to which it is bound. Pfr is the physiologically active form of the protein; therefore, exposure to red light yields physiological activity. Exposure to far-red light inhibits phytochrome activity. Together, the two forms represent the phytochrome system ([link]).

The phytochrome system acts as a biological light switch. It monitors the level, intensity, duration, and color of environmental light. The effect of red light is reversible by immediately shining far-red light on the sample, which converts the chromoprotein to the inactive Pr form. Additionally, Pfr can slowly revert to Pr in the dark, or break down over time. In all instances, the physiological response induced by red light is reversed. The active form of phytochrome (Pfr) can directly activate other molecules in the cytoplasm, or it can be trafficked to the nucleus, where it directly activates or represses specific gene expression.

Once the phytochrome system evolved, plants adapted it to serve a variety of needs. Unfiltered, full sunlight contains much more red light than far-red light. Because chlorophyll absorbs strongly in the red region of the visible spectrum, but not in the far-red region, any plant in the shade of another plant on the forest floor will be exposed to red-depleted, far-red-enriched light. The preponderance of far-red light converts phytochrome in the shaded leaves to the Pr (inactive) form, slowing growth. The nearest non-shaded (or even less-shaded) areas on the forest floor have more red light; leaves exposed to these areas sense the red light, which activates the Pfr form and induces growth. In short, plant shoots use the phytochrome system to grow away from shade and towards light. Because competition for light is so fierce in a dense plant community, the evolutionary advantages of the phytochrome system are obvious.

In seeds, the phytochrome system is not used to determine direction and quality of light (shaded versus unshaded). Instead, is it used merely to determine if there is any light at all. This is especially important in species with very small seeds, such as lettuce. Because of their size, lettuce seeds have few food reserves. Their seedlings cannot grow for long before they run out of fuel. If they germinated even a centimeter under the soil surface, the seedling would never make it into the sunlight and would die. In the dark, phytochrome is in the Pr (inactive form) and the seed will not germinate; it will only germinate if exposed to light at the surface of the soil. Upon exposure to light, Pr is converted to Pfr and germination proceeds.



The biologically inactive form of phytochrome (Pr) is converted to the biologically active form Pfr under illumination with red light. Far-red light and darkness convert the molecule back to the inactive form.

Plants also use the phytochrome system to sense the change of season. Photoperiodism is a biological response to the timing and duration of day and night. It controls flowering, setting of winter buds, and vegetative growth. Detection of seasonal changes is crucial to plant survival. Although temperature and light intensity influence plant growth, they are not reliable indicators of season because they may vary from one year to the next. Day length is a better indicator of the time of year.

As stated above, unfiltered sunlight is rich in red light but deficient in far-red light. Therefore, at dawn, all the phytochrome molecules in a leaf quickly convert to the active Pfr form, and remain in that form until sunset. In the dark, the Pfr form takes hours to slowly revert back to the Pr form. If the night is long (as in winter), all of the Pfr form reverts. If the night is short (as in summer), a considerable amount of Pfr may remain at sunrise. By sensing the Pr/Pfr ratio at dawn, a plant can determine the length of the day/night cycle. In addition, leaves retain that information for several days, allowing a comparison between the length of the previous night and the preceding several nights. Shorter nights indicate springtime to the plant; when the nights become longer, autumn is approaching. This information, along with sensing temperature and water availability, allows plants to determine the time of the year and adjust their physiology accordingly. Short-day (long-night) plants use this information to flower in the late

summer and early fall, when nights exceed a critical length (often eight or fewer hours). Long-day (short-night) plants flower during the spring, when darkness is less than a critical length (often eight to 15 hours). Not all plants use the phytochrome system in this way. Flowering in day-neutral plants is not regulated by daylength.

Note:
Career Connection
Horticulturalist
The word “horticulturist” comes from the Latin words for garden (*hortus*) and culture (*cultura*). This career has been revolutionized by progress made in the understanding of plant responses to environmental stimuli. Growers of crops, fruit, vegetables, and flowers were previously constrained by having to time their sowing and harvesting according to the season. Now, horticulturists can manipulate plants to increase leaf, flower, or fruit production by understanding how environmental factors affect plant growth and development.
Greenhouse management is an essential component of a horticulturist’s education. To lengthen the night, plants are covered with a blackout shade cloth. Long-day plants are irradiated with red light in winter to promote early flowering. For example, fluorescent (cool white) light high in blue wavelengths encourages leafy growth and is excellent for starting seedlings. Incandescent lamps (standard light bulbs) are rich in red light, and promote flowering in some plants. The timing of fruit ripening can be increased or delayed by applying plant hormones. Recently, considerable progress has been made in the development of plant breeds that are suited to different climates and resistant to pests and transportation damage. Both crop yield and quality have increased as a result of practical applications of the knowledge of plant responses to external stimuli and hormones.
Horticulturists find employment in private and governmental laboratories, greenhouses, botanical gardens, and in the production or research fields. They improve crops by applying their knowledge of genetics and plant physiology. To prepare for a horticulture career, students take classes in botany, plant physiology, plant pathology, landscape design, and plant

breeding. To complement these traditional courses, horticulture majors add studies in economics, business, computer science, and communications.

The Blue Light Responses

Phototropism—the directional bending of a plant toward or away from a light source—is a response to blue wavelengths of light. Positive phototropism is growth towards a light source ([link]), while negative phototropism (also called skototropism) is growth away from light.

The aptly-named **phototropins** are protein-based receptors responsible for mediating the phototropic response. Like all plant photoreceptors, phototropins consist of a protein portion and a light-absorbing portion, called the chromophore. In phototropins, the chromophore is a covalently-bound molecule of flavin; hence, phototropins belong to a class of proteins called flavoproteins.

Other responses under the control of phototropins are leaf opening and closing, chloroplast movement, and the opening of stomata. However, of all responses controlled by phototropins, phototropism has been studied the longest and is the best understood.

In their 1880 treatise *The Power of Movements in Plants*, Charles Darwin and his son Francis first described phototropism as the bending of seedlings toward light. Darwin observed that light was perceived by the tip of the plant (the apical meristem), but that the response (bending) took place in a different part of the plant. They concluded that the signal had to travel from the apical meristem to the base of the plant.

Azure bluets (*Houstonia caerulea*) display a phototropic response by bending toward the light. (credit: Cory Zanker)

In 1913, Peter Boysen-Jensen demonstrated that a chemical signal produced in the plant tip was responsible for the bending at the base. He cut off the tip of a seedling, covered the cut section with a layer of gelatin, and then replaced the tip. The seedling bent toward the light when illuminated. However, when impermeable mica flakes were inserted between the tip and the cut base, the seedling did not bend. A refinement of the experiment showed that the signal traveled on the shaded side of the seedling. When the mica plate was inserted on the illuminated side, the plant did bend towards the light. Therefore, the chemical signal was a growth stimulant because the phototropic response involved faster cell elongation on the shaded side than on the illuminated side. We now know that as light passes through a plant stem, it is diffracted and generates phototropin activation across the stem. Most activation occurs on the lit side, causing the plant hormone indole acetic acid (IAA) to accumulate on the shaded side. Stem cells elongate under influence of IAA.

Cryptochromes are another class of blue-light absorbing photoreceptors that also contain a flavin-based chromophore. Cryptochromes set the plants

24-hour activity cycle, also know as its circadian rhythem, using blue light cues. There is some evidence that cryptochromes work together with phototropins to mediate the phototropic response.

Note:
Link to Learning

Use the navigation menu in the left panel of this website to view images of plants in motion.

Plant Responses to Gravity

Whether or not they germinate in the light or in total darkness, shoots usually sprout up from the ground, and roots grow downward into the ground. A plant laid on its side in the dark will send shoots upward when given enough time. Gravitropism ensures that roots grow into the soil and that shoots grow toward sunlight. Growth of the shoot apical tip upward is called **negative gravitropism**, whereas growth of the roots downward is called **positive gravitropism**.

Amyloplasts (also known as **statoliths**) are specialized plastids that contain starch granules and settle downward in response to gravity. Amyloplasts are found in shoots and in specialized cells of the root cap. When a plant is tilted, the statoliths drop to the new bottom cell wall. A few hours later, the shoot or root will show growth in the new vertical direction.

The mechanism that mediates gravitropism is reasonably well understood. When amyloplasts settle to the bottom of the gravity-sensing cells in the

root or shoot, they physically contact the endoplasmic reticulum (ER), causing the release of calcium ions from inside the ER. This calcium signaling in the cells causes polar transport of the plant hormone IAA to the bottom of the cell. In roots, a high concentration of IAA inhibits cell elongation. The effect slows growth on the lower side of the root, while cells develop normally on the upper side. IAA has the opposite effect in shoots, where a higher concentration at the lower side of the shoot stimulates cell expansion, causing the shoot to grow up. After the shoot or root begin to grow vertically, the amyloplasts return to their normal position. Other hypotheses—involving the entire cell in the gravitropism effect—have been proposed to explain why some mutants that lack amyloplasts may still exhibit a weak gravitropic response.

Growth Responses

A plant's sensory response to external stimuli relies on chemical messengers (hormones). Plant hormones affect all aspects of plant life, from flowering to fruit setting and maturation, and from phototropism to leaf fall. Potentially every cell in a plant can produce plant hormones. They can act in their cell of origin or be transported to other portions of the plant body, with many plant responses involving the synergistic or antagonistic interaction of two or more hormones. In contrast, animal hormones are produced in specific glands and transported to a distant site for action, and they act alone.

Plant hormones are a group of unrelated chemical substances that affect plant morphogenesis. Five major plant hormones are traditionally described: auxins (particularly IAA), cytokinins, gibberellins, ethylene, and abscisic acid. In addition, other nutrients and environmental conditions can be characterized as growth factors.

Auxins

The term auxin is derived from the Greek word *auxein*, which means "to grow." **Auxins** are the main hormones responsible for cell elongation in

phototropism and gravitropism. They also control the differentiation of meristem into vascular tissue, and promote leaf development and arrangement. While many synthetic auxins are used as herbicides, IAA is the only naturally occurring auxin that shows physiological activity. Apical dominance—the inhibition of lateral bud formation—is triggered by auxins produced in the apical meristem. Flowering, fruit setting and ripening, and inhibition of **abscission** (leaf falling) are other plant responses under the direct or indirect control of auxins. Auxins also act as a relay for the effects of the blue light and red/far-red responses.

Commercial use of auxins is widespread in plant nurseries and for crop production. IAA is used as a rooting hormone to promote growth of adventitious roots on cuttings and detached leaves. Applying synthetic auxins to tomato plants in greenhouses promotes normal fruit development. Outdoor application of auxin promotes synchronization of fruit setting and dropping to coordinate the harvesting season. Fruits such as seedless cucumbers can be induced to set fruit by treating unfertilized plant flowers with auxins.

Cytokinins

The effect of cytokinins was first reported when it was found that adding the liquid endosperm of coconuts to developing plant embryos in culture stimulated their growth. The stimulating growth factor was found to be **cytokinin**, a hormone that promotes cytokinesis (cell division). Almost 200 naturally occurring or synthetic cytokinins are known to date. Cytokinins are most abundant in growing tissues, such as roots, embryos, and fruits, where cell division is occurring. Cytokinins are known to delay senescence in leaf tissues, promote mitosis, and stimulate differentiation of the meristem in shoots and roots. Many effects on plant development are under the influence of cytokinins, either in conjunction with auxin or another hormone. For example, apical dominance seems to result from a balance between auxins that inhibit lateral buds, and cytokinins that promote bushier growth.

Gibberellins

Gibberellins (GAs) are a group of about 125 closely related plant hormones that stimulate shoot elongation, seed germination, and fruit and flower maturation. GAs are synthesized in the root and stem apical meristems, young leaves, and seed embryos. In urban areas, GA antagonists are sometimes applied to trees under power lines to control growth and reduce the frequency of pruning.

GAs break dormancy (a state of inhibited growth and development) in the seeds of plants that require exposure to cold or light to germinate. Abscisic acid is a strong antagonist of GA action. Other effects of GAs include gender expression, seedless fruit development, and the delay of senescence in leaves and fruit. Seedless grapes are obtained through standard breeding methods and contain inconspicuous seeds that fail to develop. Because GAs are produced by the seeds, and because fruit development and stem elongation are under GA control, these varieties of grapes would normally produce small fruit in compact clusters. Maturing grapes are routinely treated with GA to promote larger fruit size, as well as looser bunches (longer stems), which reduces the instance of mildew infection ([link]).

In grapes, application of gibberellic acid increases the size of fruit and loosens clustering. (credit: Bob Nichols, USDA)

Abscisic Acid

The plant hormone **abscisic acid** (ABA) was first discovered as the agent that causes the abscission or dropping of cotton bolls. However, more recent studies indicate that ABA plays only a minor role in the abscission process. ABA accumulates as a response to stressful environmental conditions, such as dehydration, cold temperatures, or shortened day lengths. Its activity counters many of the growth-promoting effects of GAs and auxins. ABA inhibits stem elongation and induces dormancy in lateral buds.

ABA induces dormancy in seeds by blocking germination and promoting the synthesis of storage proteins. Plants adapted to temperate climates require a long period of cold temperature before seeds germinate. This mechanism protects young plants from sprouting too early during unseasonably warm weather in winter. As the hormone gradually breaks down over winter, the seed is released from dormancy and germinates when conditions are favorable in spring. Another effect of ABA is to promote the development of winter buds; it mediates the conversion of the apical meristem into a dormant bud. Low soil moisture causes an increase in ABA, which causes stomata to close, reducing water loss in winter buds.

Ethylene

Ethylene is associated with fruit ripening, flower wilting, and leaf fall. Ethylene is unusual because it is a volatile gas (C_2H_4). Hundreds of years ago, when gas street lamps were installed in city streets, trees that grew close to lamp posts developed twisted, thickened trunks and shed their leaves earlier than expected. These effects were caused by ethylene volatilizing from the lamps.

Aging tissues (especially senescing leaves) and nodes of stems produce ethylene. The best-known effect of the hormone, however, is the promotion of fruit ripening. Ethylene stimulates the conversion of starch and acids to sugars. Some people store unripe fruit, such as avocadoes, in a sealed paper bag to accelerate ripening; the gas released by the first fruit to mature will speed up the maturation of the remaining fruit. Ethylene also triggers leaf and fruit abscission, flower fading and dropping, and promotes germination in some cereals and sprouting of bulbs and potatoes.

Ethylene is widely used in agriculture. Commercial fruit growers control the timing of fruit ripening with application of the gas. Horticulturalists inhibit leaf dropping in ornamental plants by removing ethylene from greenhouses using fans and ventilation.

Nontraditional Hormones

Recent research has discovered a number of compounds that also influence plant development. Their roles are less understood than the effects of the major hormones described so far.

Jasmonates play a major role in defense responses to herbivory. Their levels increase when a plant is wounded by a predator, resulting in an increase in toxic secondary metabolites. They contribute to the production of volatile compounds that attract natural enemies of predators. For example, chewing of tomato plants by caterpillars leads to an increase in jasmonic acid levels, which in turn triggers the release of volatile compounds that attract predators of the pest.

Oligosaccharins also play a role in plant defense against bacterial and fungal infections. They act locally at the site of injury, and can also be transported to other tissues. **Strigolactones** promote seed germination in some species and inhibit lateral apical development in the absence of auxins. Strigolactones also play a role in the establishment of mycorrhizae, a mutualistic association of plant roots and fungi. Brassinosteroids are important to many developmental and physiological processes. Signals between these compounds and other hormones, notably auxin and GAs, amplifies their physiological effect. Apical dominance, seed germination, gravitropism, and resistance to freezing are all positively influenced by hormones. Root growth and fruit dropping are inhibited by steroids.

Plant Responses to Wind and Touch

The shoot of a pea plant winds around a trellis, while a tree grows on an angle in response to strong prevailing winds. These are examples of how plants respond to touch or wind.

The movement of a plant subjected to constant directional pressure is called **thigmotropism**, from the Greek words *thigma* meaning “touch,” and *tropism* implying “direction.” Tendrils are one example of this. The meristematic region of tendrils is very touch sensitive; light touch will evoke a quick coiling response. Cells in contact with a support surface

contract, whereas cells on the opposite side of the support expand ([link]). Application of jasmonic acid is sufficient to trigger tendril coiling without a mechanical stimulus.

A **thigmonastic** response is a touch response independent of the direction of stimulus [link]. In the Venus flytrap, two modified leaves are joined at a hinge and lined with thin fork-like tines along the outer edges. Tiny hairs are located inside the trap. When an insect brushes against these trigger hairs, touching two or more of them in succession, the leaves close quickly, trapping the prey. Glands on the leaf surface secrete enzymes that slowly digest the insect. The released nutrients are absorbed by the leaves, which reopen for the next meal.

Thigmomorphogenesis is a slow developmental change in the shape of a plant subjected to continuous mechanical stress. When trees bend in the wind, for example, growth is usually stunted and the trunk thickens. Strengthening tissue, especially xylem, is produced to add stiffness to resist the wind’s force. Researchers hypothesize that mechanical strain induces growth and differentiation to strengthen the tissues. Ethylene and jasmonate are likely involved in thigmomorphogenesis.

Note:
Link to Learning

Use the menu at the left to navigate to three short movies: a Venus fly trap capturing prey, the progressive closing of sensitive plant leaflets, and the twining of tendrils.

Defense Responses against Herbivores and Pathogens

Plants face two types of enemies: herbivores and pathogens. Herbivores both large and small use plants as food, and actively chew them. Pathogens are agents of disease. These infectious microorganisms, such as fungi, bacteria, and nematodes, live off of the plant and damage its tissues. Plants have developed a variety of strategies to discourage or kill attackers.

The first line of defense in plants is an intact and impenetrable barrier. Bark and the waxy cuticle can protect against predators. Other adaptations against herbivory include thorns, which are modified branches, and spines, which are modified leaves. They discourage animals by causing physical damage and inducing rashes and allergic reactions. A plant’s exterior protection can be compromised by mechanical damage, which may provide an entry point for pathogens. If the first line of defense is breached, the plant must resort to a different set of defense mechanisms, such as toxins and enzymes.

Secondary metabolites are compounds that are not directly derived from photosynthesis and are not necessary for respiration or plant growth and development. Many metabolites are toxic, and can even be lethal to animals that ingest them. Some metabolites are alkaloids, which discourage predators with noxious odors (such as the volatile oils of mint and sage) or repellent tastes (like the bitterness of quinine). Other alkaloids affect herbivores by causing either excessive stimulation (caffeine is one example) or the lethargy associated with opioids. Some compounds become toxic after ingestion; for instance, glycol cyanide in the cassava root releases cyanide only upon ingestion by the herbivore.

Mechanical wounding and predator attacks activate defense and protection mechanisms both in the damaged tissue and at sites farther from the injury location. Some defense reactions occur within minutes: others over several hours. The infected and surrounding cells may die, thereby stopping the spread of infection.

Long-distance signaling elicits a systemic response aimed at deterring the predator. As tissue is damaged, jasmonates may promote the synthesis of compounds that are toxic to predators. Jasmonates also elicit the synthesis

of volatile compounds that attract parasitoids, which are insects that spend their developing stages in or on another insect, and eventually kill their host. The plant may activate abscission of injured tissue if it is damaged beyond repair.

Section Summary

Plants respond to light by changes in morphology and activity. Irradiation by red light converts the photoreceptor phytochrome to its far-red light-absorbing form—Pfr. This form controls germination and flowering in response to length of day, as well as triggers photosynthesis in dormant plants or those that just emerged from the soil. Blue-light receptors, cryptochromes, and phototropins are responsible for phototropism. Amyloplasts, which contain heavy starch granules, sense gravity. Shoots exhibit negative gravitropism, whereas roots exhibit positive gravitropism. Plant hormones—naturally occurring compounds synthesized in small amounts—can act both in the cells that produce them and in distant tissues and organs. Auxins are responsible for apical dominance, root growth, directional growth toward light, and many other growth responses. Cytokinins stimulate cell division and counter apical dominance in shoots. Gibberellins inhibit dormancy of seeds and promote stem growth. Abscisic acid induces dormancy in seeds and buds, and protects plants from excessive water loss by promoting stomatal closure. Ethylene gas speeds up fruit ripening and dropping of leaves. Plants respond to touch by rapid movements (thigmotropy and thigmonasty) and slow differential growth (thigmomorphogenesis). Plants have evolved defense mechanisms against predators and pathogens. Physical barriers like bark and spines protect tender tissues. Plants also have chemical defenses, including toxic secondary metabolites and hormones, which elicit additional defense mechanisms.

Review Questions

Exercise:

Problem:

The main photoreceptor that triggers phototropism is a ________.

a. phytochrome
b. cryptochrome
c. phototropin
d. carotenoid

Solution:

C

Exercise:

Problem: Phytochrome is a plant pigment protein that:

a. mediates plant infection
b. promotes plant growth
c. mediates morphological changes in response to red and far-red light
d. inhibits plant growth

Solution:

C

Exercise:

Problem:

A mutant plant has roots that grow in all directions. Which of the following organelles would you expect to be missing in the cell?

a. mitochondria
b. amyloplast
c. chloroplast

d. nucleus

Solution:

B

Exercise:

Problem:

After buying green bananas or unripe avocadoes, they can be kept in a brown bag to ripen. The hormone released by the fruit and trapped in the bag is probably:

a. abscisic acid
b. cytokinin
c. ethylene
d. gibberellic acid

Solution:

C

Exercise:

Problem:

A decrease in the level of which hormone releases seeds from dormancy?

a. abscisic acid
b. cytokinin
c. ethylene
d. gibberellic acid

Solution:

A

Exercise:

Problem:

A seedling germinating under a stone grows at an angle away from the stone and upward. This response to touch is called _________.

a. gravitropism
b. thigmonasty
c. thigmotropism
d. skototropism

Solution:

C

Free Response

Exercise:

Problem:

Owners and managers of plant nurseries have to plan lighting schedules for a long-day plant that will flower in February. What lighting periods will be most effective? What color of light should be chosen?

Solution:

A long-day plant needs a higher proportion of the Pfr form to Pr form of phytochrome. The plant requires long periods of illumination with light enriched in the red range of the spectrum.

Exercise:

Problem:

What are the major benefits of gravitropism for a germinating seedling?

Solution:

Gravitropism will allow roots to dig deep into the soil to find water and minerals, whereas the seedling will grow towards light to enable photosynthesis.

Exercise:

Problem:

Fruit and vegetable storage facilities are usually refrigerated and well ventilated. Why are these conditions advantageous?

Solution:

Refrigeration slows chemical reactions, including fruit maturation. Ventilation removes the ethylene gas that speeds up fruit ripening.

Exercise:

Problem:

Stomata close in response to bacterial infection. Why is this response a mechanism of defense for the plant? Which hormone is most likely to mediate this response?

Solution:

To prevent further entry of pathogens, stomata close, even if they restrict entry of CO_2. Some pathogens secrete virulence factors that inhibit the closing of stomata. Abscisic acid is the stress hormone responsible for inducing closing of stomata.

Glossary

abscisic acid (ABA)
: plant hormone that induces dormancy in seeds and other organs

abscission
: physiological process that leads to the fall of a plant organ (such as leaf or petal drop)

auxin
: plant hormone that influences cell elongation (in phototropism), gravitropism, apical dominance and root growth

chromophore
: molecule that absorbs light

cryptochrome
: protein that absorbs light in the blue and ultraviolet regions of the light spectrum

cytokinin
: plant hormone that promotes cell division

ethylene
: volatile plant hormone that is associated with fruit ripening, flower wilting, and leaf fall

gibberellin (GA)
: plant hormone that stimulates shoot elongation, seed germination, and the maturation and dropping of fruit and flowers

jasmonates
: small family of compounds derived from the fatty acid linoleic acid

negative gravitropism
: growth away from Earth’s gravity

oligosaccharin
: hormone important in plant defenses against bacterial and fungal infections

photomorphogenesis
    growth and development of plants in response to light

photoperiodism
    occurrence of plant processes, such as germination and flowering, according to the time of year

phototropin
    blue-light receptor that promotes phototropism, stomatal opening and closing, and other responses that promote photosynthesis

phototropism
    directional bending of a plant toward a light source

phytochrome
    plant pigment protein that exists in two reversible forms (Pr and Pfr) and mediates morphologic changes in response to red light

positive gravitropism
    growth toward Earth's gravitational center

statolith
    (also, **amyloplast**) plant organelle that contains heavy starch granules

strigolactone
    hormone that promotes seed germination in some species and inhibits lateral apical development in the absence of auxins

thigmomorphogenesis
    developmental response to touch

thigmonastic
    directional growth of a plant independent of the direction in which contact is applied

thigmotropism
    directional growth of a plant in response to constant contact

Pollination and Fertilization
By the end of this section, you will be able to:

- Describe what must occur for plant fertilization
- Explain cross-pollination and the ways in which it takes place
- Describe the process that leads to the development of a seed
- Define double fertilization

In angiosperms, **pollination** is defined as the placement or transfer of pollen from the anther to the stigma of the same flower or another flower. In gymnosperms, pollination involves pollen transfer from the male cone to the female cone. Upon transfer, the pollen germinates to form the pollen tube and the sperm for fertilizing the egg. Pollination has been well studied since the time of Gregor Mendel. Mendel successfully carried out self- as well as cross-pollination in garden peas while studying how characteristics were passed on from one generation to the next. Today’s crops are a result of plant breeding, which employs artificial selection to produce the present-day cultivars. A case in point is today's corn, which is a result of years of breeding that started with its ancestor, teosinte. The teosinte that the ancient Mayans originally began cultivating had tiny seeds—vastly different from today’s relatively giant ears of corn. Interestingly, though these two plants appear to be entirely different, the genetic difference between them is miniscule.

Pollination takes two forms: self-pollination and cross-pollination. **Self-pollination** occurs when the pollen from the anther is deposited on the stigma of the same flower, or another flower on the same plant. **Cross-pollination** is the transfer of pollen from the anther of one flower to the stigma of another flower on a different individual of the same species. Self-pollination occurs in flowers where the stamen and carpel mature at the same time, and are positioned so that the pollen can land on the flower’s stigma. This method of pollination does not require an investment from the plant to provide nectar and pollen as food for pollinators.

Note:
Link to Learning

Explore this interactive website to review self-pollination and cross-pollination.

Living species are designed to ensure survival of their progeny; those that fail become extinct. Genetic diversity is therefore required so that in changing environmental or stress conditions, some of the progeny can survive. Self-pollination leads to the production of plants with less genetic diversity, since genetic material from the same plant is used to form gametes, and eventually, the zygote. In contrast, cross-pollination—or out-crossing—leads to greater genetic diversity because the microgametophyte and megagametophyte are derived from different plants.

Because cross-pollination allows for more genetic diversity, plants have developed many ways to avoid self-pollination. In some species, the pollen and the ovary mature at different times. These flowers make self-pollination nearly impossible. By the time pollen matures and has been shed, the stigma of this flower is mature and can only be pollinated by pollen from another flower. Some flowers have developed physical features that prevent self-pollination. The primrose is one such flower. Primroses have evolved two flower types with differences in anther and stigma length: the pin-eyed flower has anthers positioned at the pollen tube's halfway point, and the thrum-eyed flower's stigma is likewise located at the halfway point. Insects easily cross-pollinate while seeking the nectar at the bottom of the pollen tube. This phenomenon is also known as heterostyly. Many plants, such as cucumber, have male and female flowers located on different parts of the plant, thus making self-pollination difficult. In yet other species, the male and female flowers are borne on different plants (dioecious). All of these are barriers to self-pollination; therefore, the plants depend on pollinators to transfer pollen. The majority of pollinators are biotic agents such as insects

(like bees, flies, and butterflies), bats, birds, and other animals. Other plant species are pollinated by abiotic agents, such as wind and water.

Note:
Everyday Connection
Incompatibility Genes in Flowers
In recent decades, incompatibility genes—which prevent pollen from germinating or growing into the stigma of a flower—have been discovered in many angiosperm species. If plants do not have compatible genes, the pollen tube stops growing. Self-incompatibility is controlled by the S (sterility) locus. Pollen tubes have to grow through the tissue of the stigma and style before they can enter the ovule. The carpel is selective in the type of pollen it allows to grow inside. The interaction is primarily between the pollen and the stigma epidermal cells. In some plants, like cabbage, the pollen is rejected at the surface of the stigma, and the unwanted pollen does not germinate. In other plants, pollen tube germination is arrested after growing one-third the length of the style, leading to pollen tube death. Pollen tube death is due either to apoptosis (programmed cell death) or to degradation of pollen tube RNA. The degradation results from the activity of a ribonuclease encoded by the S locus. The ribonuclease is secreted from the cells of the style in the extracellular matrix, which lies alongside the growing pollen tube.
In summary, self-incompatibility is a mechanism that prevents self-fertilization in many flowering plant species. The working of this self-incompatibility mechanism has important consequences for plant breeders because it inhibits the production of inbred and hybrid plants.

Pollination by Insects

Bees are perhaps the most important pollinator of many garden plants and most commercial fruit trees ([link]). The most common species of bees are bumblebees and honeybees. Since bees cannot see the color red, bee-pollinated flowers usually have shades of blue, yellow, or other colors. Bees collect energy-rich pollen or nectar for their survival and energy needs.

They visit flowers that are open during the day, are brightly colored, have a strong aroma or scent, and have a tubular shape, typically with the presence of a nectar guide. A **nectar guide** includes regions on the flower petals that are visible only to bees, and not to humans; it helps to guide bees to the center of the flower, thus making the pollination process more efficient. The pollen sticks to the bees' fuzzy hair, and when the bee visits another flower, some of the pollen is transferred to the second flower. Recently, there have been many reports about the declining population of honeybees. Many flowers will remain unpollinated and not bear seed if honeybees disappear. The impact on commercial fruit growers could be devastating.

Insects, such as bees, are important agents of pollination. (credit: modification of work by Jon Sullivan)

Many flies are attracted to flowers that have a decaying smell or an odor of rotting flesh. These flowers, which produce nectar, usually have dull colors, such as brown or purple. They are found on the corpse flower or voodoo lily (*Amorphophallus*), dragon arum (*Dracunculus*), and carrion flower (*Stapleia*, *Rafflesia*). The nectar provides energy, whereas the pollen

provides protein. Wasps are also important insect pollinators, and pollinate many species of figs.

Butterflies, such as the monarch, pollinate many garden flowers and wildflowers, which usually occur in clusters. These flowers are brightly colored, have a strong fragrance, are open during the day, and have nectar guides to make access to nectar easier. The pollen is picked up and carried on the butterfly’s limbs. Moths, on the other hand, pollinate flowers during the late afternoon and night. The flowers pollinated by moths are pale or white and are flat, enabling the moths to land. One well-studied example of a moth-pollinated plant is the yucca plant, which is pollinated by the yucca moth. The shape of the flower and moth have adapted in such a way as to allow successful pollination. The moth deposits pollen on the sticky stigma for fertilization to occur later. The female moth also deposits eggs into the ovary. As the eggs develop into larvae, they obtain food from the flower and developing seeds. Thus, both the insect and flower benefit from each other in this symbiotic relationship. The corn earworm moth and Gaura plant have a similar relationship ([link]).

A corn earworm sips nectar from a night-blooming Gaura plant. (credit: Juan Lopez, USDA ARS)

Pollination by Bats

In the tropics and deserts, bats are often the pollinators of nocturnal flowers such as agave, guava, and morning glory. The flowers are usually large and white or pale-colored; thus, they can be distinguished from the dark surroundings at night. The flowers have a strong, fruity, or musky fragrance and produce large amounts of nectar. They are naturally large and wide-mouthed to accommodate the head of the bat. As the bats seek the nectar, their faces and heads become covered with pollen, which is then transferred to the next flower.

Pollination by Birds

Many species of small birds, such as the hummingbird ([link]) and sun birds, are pollinators for plants such as orchids and other wildflowers. Flowers visited by birds are usually sturdy and are oriented in such a way as to allow the birds to stay near the flower without getting their wings entangled in the nearby flowers. The flower typically has a curved, tubular shape, which allows access for the bird's beak. Brightly colored, odorless flowers that are open during the day are pollinated by birds. As a bird seeks energy-rich nectar, pollen is deposited on the bird's head and neck and is then transferred to the next flower it visits. Botanists have been known to determine the range of extinct plants by collecting and identifying pollen from 200-year-old bird specimens from the same site.

Hummingbirds have adaptations that allow them to reach the nectar of certain tubular flowers. (credit: Lori Branham)

Pollination by Wind

Most species of conifers, and many angiosperms, such as grasses, maples and oaks, are pollinated by wind. Pine cones are brown and unscented, while the flowers of wind-pollinated angiosperm species are usually green, small, may have small or no petals, and produce large amounts of pollen. Unlike the typical insect-pollinated flowers, flowers adapted to pollination by wind do not produce nectar or scent. In wind-pollinated species, the microsporangia hang out of the flower, and, as the wind blows, the lightweight pollen is carried with it ([link]). The flowers usually emerge early in the spring, before the leaves, so that the leaves do not block the movement of the wind. The pollen is deposited on the exposed feathery stigma of the flower ([link]).

A person knocks pollen from a pine tree.

(a) (b)

These male (a) and female (b) catkins are from the goat willow tree (*Salix caprea*). Note how both structures are light and feathery to better disperse and catch the wind-blown pollen.

Pollination by Water

Some weeds, such as Australian sea grass and pond weeds, are pollinated by water. The pollen floats on water, and when it comes into contact with the flower, it is deposited inside the flower.

Note:
Evolution Connection
Pollination by Deception
Orchids are highly valued flowers, with many rare varieties ([link]). They grow in a range of specific habitats, mainly in the tropics of Asia, South America, and Central America. At least 25,000 species of orchids have been identified.

Certain orchids use food deception or sexual deception to attract pollinators. Shown here is a bee orchid (*Ophrys apifera*). (credit: David Evans)

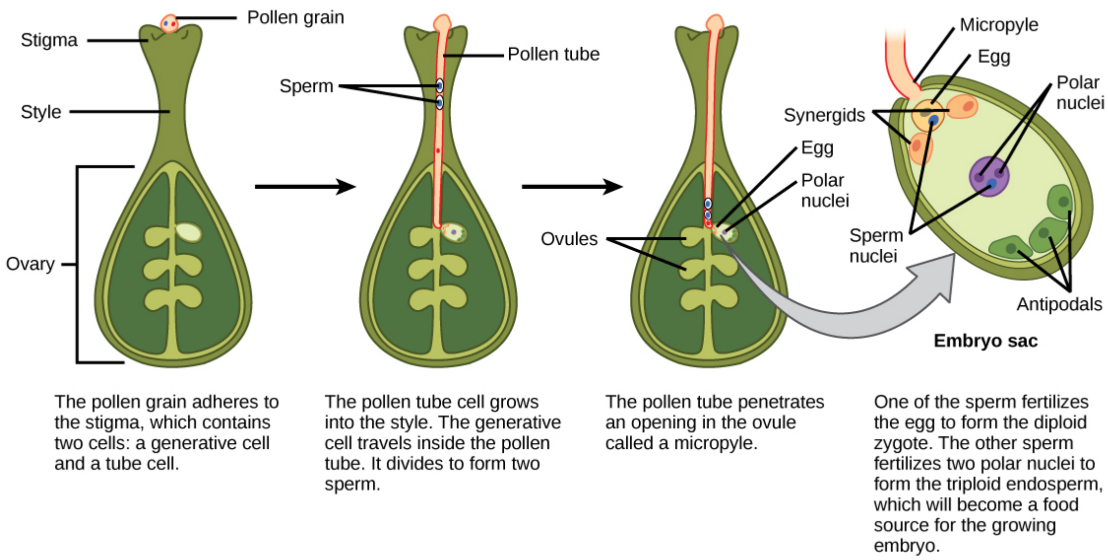
Flowers often attract pollinators with food rewards, in the form of nectar. However, some species of orchid are an exception to this standard: they have evolved different ways to attract the desired pollinators. They use a method known as food deception, in which bright colors and perfumes are offered, but no food. *Anacamptis morio*, commonly known as the green-winged orchid, bears bright purple flowers and emits a strong scent. The bumblebee, its main pollinator, is attracted to the flower because of the strong scent—which usually indicates food for a bee—and in the process, picks up the pollen to be transported to another flower.

Other orchids use sexual deception. *Chiloglottis trapeziformis* emits a compound that smells the same as the pheromone emitted by a female wasp to attract male wasps. The male wasp is attracted to the scent, lands on the orchid flower, and in the process, transfers pollen. Some orchids, like the Australian hammer orchid, use scent as well as visual trickery in yet another sexual deception strategy to attract wasps. The flower of this orchid mimics the appearance of a female wasp and emits a pheromone. The male wasp tries to mate with what appears to be a female wasp, and in the process, picks up pollen, which it then transfers to the next counterfeit mate.

Double Fertilization

After pollen is deposited on the stigma, it must germinate and grow through the style to reach the ovule. The microspores, or the pollen, contain two cells: the pollen tube cell and the generative cell. The pollen tube cell grows into a pollen tube through which the generative cell travels. The germination of the pollen tube requires water, oxygen, and certain chemical signals. As it travels through the style to reach the embryo sac, the pollen tube’s growth is supported by the tissues of the style. In the meantime, if the generative cell has not already split into two cells, it now divides to form two sperm cells. The pollen tube is guided by the chemicals secreted by the synergids present in the embryo sac, and it enters the ovule sac through the micropyle. Of the two sperm cells, one sperm fertilizes the egg cell, forming a diploid zygote; the other sperm fuses with the two polar nuclei,

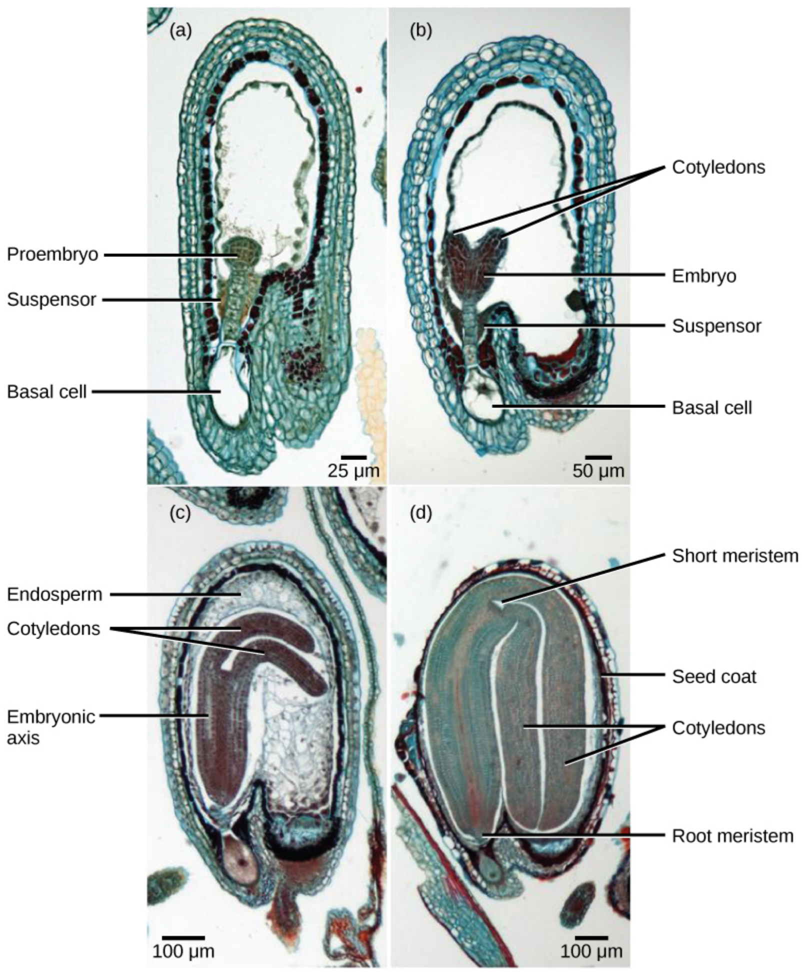
forming a triploid cell that develops into the **endosperm**. Together, these two fertilization events in angiosperms are known as **double fertilization** ([link]). After fertilization is complete, no other sperm can enter. The fertilized ovule forms the seed, whereas the tissues of the ovary become the fruit, usually enveloping the seed.



In angiosperms, one sperm fertilizes the egg to form the 2*n* zygote, and the other sperm fertilizes the central cell to form the 3*n* endosperm. This is called a double fertilization.

After fertilization, the zygote divides to form two cells: the upper cell, or terminal cell, and the lower, or basal, cell. The division of the basal cell gives rise to the **suspensor**, which eventually makes connection with the maternal tissue. The suspensor provides a route for nutrition to be transported from the mother plant to the growing embryo. The terminal cell also divides, giving rise to a globular-shaped proembryo ([link]**a**). In dicots (eudicots), the developing embryo has a heart shape, due to the presence of the two rudimentary **cotyledons** ([link]**b**). In non-endospermic dicots, such as *Capsella bursa*, the endosperm develops initially, but is then digested, and the food reserves are moved into the two cotyledons. As the embryo

and cotyledons enlarge, they run out of room inside the developing seed, and are forced to bend ([link]**c**). Ultimately, the embryo and cotyledons fill the seed ([link]**d**), and the seed is ready for dispersal. Embryonic development is suspended after some time, and growth is resumed only when the seed germinates. The developing seedling will rely on the food reserves stored in the cotyledons until the first set of leaves begin photosynthesis.



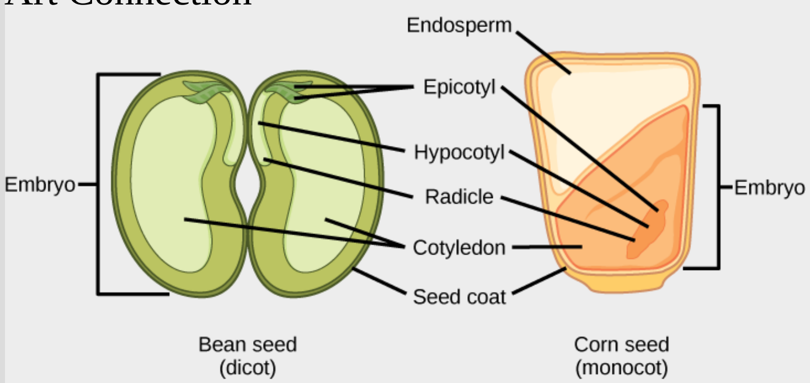
Shown are the stages of embryo development in the ovule of a shepherd's purse (*Capsella bursa*). After fertilization, the zygote divides to form an upper terminal cell and a lower basal cell. (a) In the first

stage of development, the terminal cell divides, forming a globular pro-embryo. The basal cell also divides, giving rise to the suspensor. (b) In the second stage, the developing embryo has a heart shape due to the presence of cotyledons. (c) In the third stage, the growing embryo runs out of room and starts to bend. (d) Eventually, it completely fills the seed. (credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Development of the Seed

The mature ovule develops into the seed. A typical seed contains a seed coat, cotyledons, endosperm, and a single embryo ([link]).

Note:
Art Connection



The structures of dicot and monocot seeds are shown. Dicots (left) have two cotyledons. Monocots, such as corn (right), have one cotyledon, called the scutellum; it

channels nutrition to the growing embryo. Both monocot and dicot embryos have a plumule that forms the leaves, a hypocotyl that forms the stem, and a radicle that forms the root. The embryonic axis comprises everything between the plumule and the radicle, not including the cotyledon(s).

What is of the following statements is true?

a. Both monocots and dicots have an endosperm.
b. The radicle develops into the root.
c. The plumule is part of the epicotyl
d. The endosperm is part of the embryo.

The storage of food reserves in angiosperm seeds differs between monocots and dicots. In monocots, such as corn and wheat, the single cotyledon is called a **scutellum**; the scutellum is connected directly to the embryo via vascular tissue (xylem and phloem). Food reserves are stored in the large endosperm. Upon germination, enzymes are secreted by the **aleurone**, a single layer of cells just inside the seed coat that surrounds the endosperm and embryo. The enzymes degrade the stored carbohydrates, proteins and lipids, the products of which are absorbed by the scutellum and transported via a vasculature strand to the developing embryo. Therefore, the scutellum can be seen to be an absorptive organ, not a storage organ.

The two cotyledons in the dicot seed also have vascular connections to the embryo. In **endospermic dicots**, the food reserves are stored in the endosperm. During germination, the two cotyledons therefore act as absorptive organs to take up the enzymatically released food reserves, much like in monocots (monocots, by definition, also have endospermic seeds). Tobacco (*Nicotiana tabaccum*), tomato (*Solanum lycopersicum*), and pepper (*Capsicum annuum*) are examples of endospermic dicots. In **non-endospermic dicots**, the triploid endosperm develops normally following

double fertilization, but the endosperm food reserves are quickly remobilized and moved into the developing cotyledon for storage. The two halves of a peanut seed (*Arachis hypogaea*) and the split peas (*Pisum sativum*) of split pea soup are individual cotyledons loaded with food reserves.

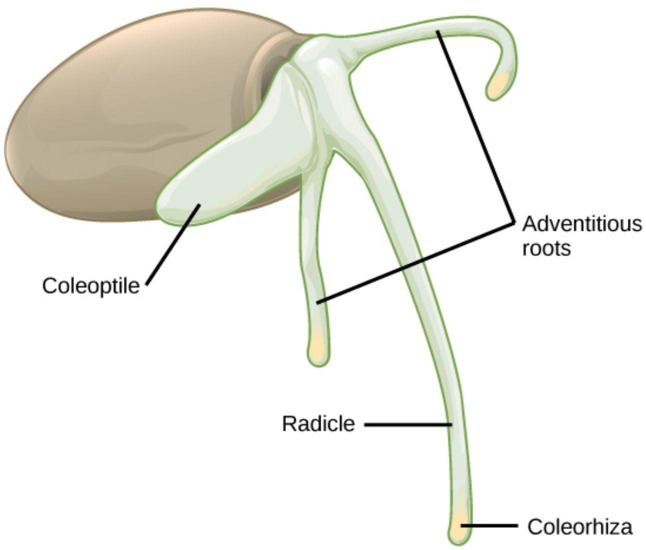
The seed, along with the ovule, is protected by a seed coat that is formed from the integuments of the ovule sac. In dicots, the seed coat is further divided into an outer coat known as the **testa** and inner coat known as the **tegmen**.

The embryonic axis consists of three parts: the plumule, the radicle, and the hypocotyl. The portion of the embryo between the cotyledon attachment point and the radicle is known as the **hypocotyl** (hypocotyl means “below the cotyledons”). The embryonic axis terminates in a **radicle** (the embryonic root), which is the region from which the root will develop. In dicots, the hypocotyls extend above ground, giving rise to the stem of the plant. In monocots, the hypocotyl does not show above ground because monocots do not exhibit stem elongation. The part of the embryonic axis that projects above the cotyledons is known as the **epicotyl**. The **plumule** is composed of the epicotyl, young leaves, and the shoot apical meristem.

Upon germination in dicot seeds, the epicotyl is shaped like a hook with the plumule pointing downwards. This shape is called the plumule hook, and it persists as long as germination proceeds in the dark. Therefore, as the epicotyl pushes through the tough and abrasive soil, the plumule is protected from damage. Upon exposure to light, the hypocotyl hook straightens out, the young foliage leaves face the sun and expand, and the epicotyl continues to elongate. During this time, the radicle is also growing and producing the primary root. As it grows downward to form the tap root, lateral roots branch off to all sides, producing the typical dicot tap root system.

In monocot seeds ([link]), the testa and tegmen of the seed coat are fused. As the seed germinates, the primary root emerges, protected by the root-tip covering: the **coleorhiza**. Next, the primary shoot emerges, protected by the **coleoptile**: the covering of the shoot tip. Upon exposure to light (i.e. when the plumule has exited the soil and the protective coleoptile is no longer

needed), elongation of the coleoptile ceases and the leaves expand and unfold. At the other end of the embryonic axis, the primary root soon dies, while other, adventitious roots (roots that do not arise from the usual place – i.e. the root) emerge from the base of the stem. This gives the monocot a fibrous root system.



As this monocot grass seed germinates, the primary root, or radicle, emerges first, followed by the primary shoot, or coleoptile, and the adventitious roots.

Seed Germination

Many mature seeds enter a period of inactivity, or extremely low metabolic activity: a process known as **dormancy**, which may last for months, years or even centuries. Dormancy helps keep seeds viable during unfavorable conditions. Upon a return to favorable conditions, seed germination takes place. Favorable conditions could be as diverse as moisture, light, cold, fire,

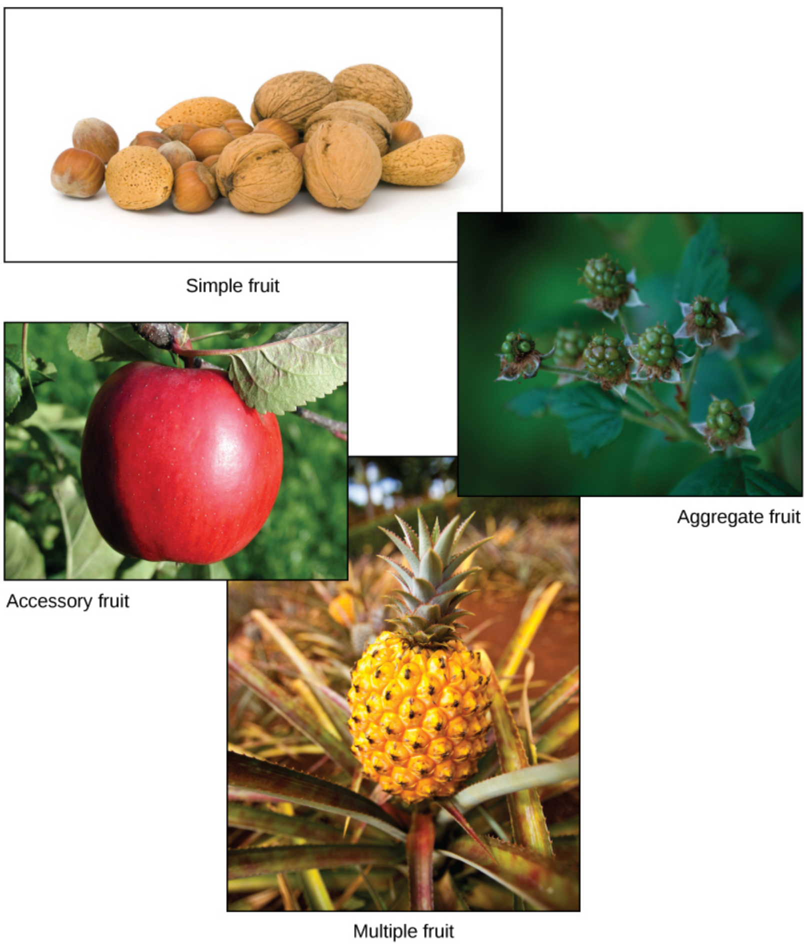
or chemical treatments. After heavy rains, many new seedlings emerge. Forest fires also lead to the emergence of new seedlings. Some seeds require **vernalization** (cold treatment) before they can germinate. This guarantees that seeds produced by plants in temperate climates will not germinate until the spring. Plants growing in hot climates may have seeds that need a heat treatment in order to germinate, to avoid germination in the hot, dry summers. In many seeds, the presence of a thick seed coat retards the ability to germinate. **Scarification**, which includes mechanical or chemical processes to soften the seed coat, is often employed before germination. Presoaking in hot water, or passing through an acid environment, such as an animal's digestive tract, may also be employed.

Depending on seed size, the time taken for a seedling to emerge may vary. Species with large seeds have enough food reserves to germinate deep below ground, and still extend their epicotyl all the way to the soil surface. Seeds of small-seeded species usually require light as a germination cue. This ensures the seeds only germinate at or near the soil surface (where the light is greatest). If they were to germinate too far underneath the surface, the developing seedling would not have enough food reserves to reach the sunlight.

Development of Fruit and Fruit Types

After fertilization, the ovary of the flower usually develops into the fruit. Fruits are usually associated with having a sweet taste; however, not all fruits are sweet. Botanically, the term "fruit" is used for a ripened ovary. In most cases, flowers in which fertilization has taken place will develop into fruits, and flowers in which fertilization has not taken place will not. Some fruits develop from the ovary and are known as true fruits, whereas others develop from other parts of the female gametophyte and are known as accessory fruits. The fruit encloses the seeds and the developing embryo, thereby providing it with protection. Fruits are of many types, depending on their origin and texture. The sweet tissue of the blackberry, the red flesh of the tomato, the shell of the peanut, and the hull of corn (the tough, thin part that gets stuck in your teeth when you eat popcorn) are all fruits. As the fruit matures, the seeds also mature.

Fruits may be classified as simple, aggregate, multiple, or accessory, depending on their origin ([link]). If the fruit develops from a single carpel or fused carpels of a single ovary, it is known as a **simple fruit**, as seen in nuts and beans. An **aggregate fruit** is one that develops from more than one carpel, but all are in the same flower: the mature carpels fuse together to form the entire fruit, as seen in the raspberry. **Multiple fruit** develops from an inflorescence or a cluster of flowers. An example is the pineapple, where the flowers fuse together to form the fruit. **Accessory fruits** (sometimes called false fruits) are not derived from the ovary, but from another part of the flower, such as the receptacle (strawberry) or the hypanthium (apples and pears).



There are four main types of fruits. Simple fruits, such as these nuts, are derived from a

single ovary. Aggregate fruits, like raspberries, form from many carpels that fuse together. Multiple fruits, such as pineapple, form from a cluster of flowers called an inflorescence. Accessory fruit, like the apple, are formed from a part of the plant other than the ovary. (credit "nuts": modification of work by Petr Kratochvil; credit "raspberries": modification of work by Cory Zanker; credit "pineapple": modification of work by Howie Le; credit "apple": modification of work by Paolo Neo)

Fruits generally have three parts: the **exocarp** (the outermost skin or covering), the **mesocarp** (middle part of the fruit), and the **endocarp** (the inner part of the fruit). Together, all three are known as the **pericarp**. The mesocarp is usually the fleshy, edible part of the fruit; however, in some fruits, such as the almond, the endocarp is the edible part. In many fruits, two or all three of the layers are fused, and are indistinguishable at maturity. Fruits can be dry or fleshy. Furthermore, fruits can be divided into dehiscent or indehiscent types. Dehiscent fruits, such as peas, readily release their seeds, while indehiscent fruits, like peaches, rely on decay to release their seeds.

Fruit and Seed Dispersal

The fruit has a single purpose: seed dispersal. Seeds contained within fruits need to be dispersed far from the mother plant, so they may find favorable and less competitive conditions in which to germinate and grow.

Some fruit have built-in mechanisms so they can disperse by themselves, whereas others require the help of agents like wind, water, and animals ([link]). Modifications in seed structure, composition, and size help in dispersal. Wind-dispersed fruit are lightweight and may have wing-like

appendages that allow them to be carried by the wind. Some have a parachute-like structure to keep them afloat. Some fruits—for example, the dandelion—have hairy, weightless structures that are suited to dispersal by wind.

Seeds dispersed by water are contained in light and buoyant fruit, giving them the ability to float. Coconuts are well known for their ability to float on water to reach land where they can germinate. Similarly, willow and silver birches produce lightweight fruit that can float on water.

Animals and birds eat fruits, and the seeds that are not digested are excreted in their droppings some distance away. Some animals, like squirrels, bury seed-containing fruits for later use; if the squirrel does not find its stash of fruit, and if conditions are favorable, the seeds germinate. Some fruits, like the cocklebur, have hooks or sticky structures that stick to an animal's coat and are then transported to another place. Humans also play a big role in dispersing seeds when they carry fruits to new places and throw away the inedible part that contains the seeds.

All of the above mechanisms allow for seeds to be dispersed through space, much like an animal's offspring can move to a new location. Seed dormancy, which was described earlier, allows plants to disperse their progeny through time: something animals cannot do. Dormant seeds can wait months, years, or even decades for the proper conditions for germination and propagation of the species.

Fruits and seeds are dispersed by various means. (a) Dandelion seeds are dispersed by wind, the (b) coconut seed is dispersed by water, and the (c) acorn is dispersed by

animals that cache and then forget it. (credit a: modification of work by "Rosendahl"/Flickr; credit b: modification of work by Shine Oa; credit c: modification of work by Paolo Neo)

Section Summary

For fertilization to occur in angiosperms, pollen has to be transferred to the stigma of a flower: a process known as pollination. Gymnosperm pollination involves the transfer of pollen from a male cone to a female cone. When the pollen of the flower is transferred to the stigma of the same flower, it is called self-pollination. Cross-pollination occurs when pollen is transferred from one flower to another flower on the same plant, or another plant. Cross-pollination requires pollinating agents such as water, wind, or animals, and increases genetic diversity. After the pollen lands on the stigma, the tube cell gives rise to the pollen tube, through which the generative nucleus migrates. The pollen tube gains entry through the micropyle on the ovule sac. The generative cell divides to form two sperm cells: one fuses with the egg to form the diploid zygote, and the other fuses with the polar nuclei to form the endosperm, which is triploid in nature. This is known as double fertilization. After fertilization, the zygote divides to form the embryo and the fertilized ovule forms the seed. The walls of the ovary form the fruit in which the seeds develop. The seed, when mature, will germinate under favorable conditions and give rise to the diploid sporophyte.

Art Connections

Exercise:

Problem: [link] What is the function of the cotyledon?

a. It develops into the root.

b. It provides nutrition for the embryo.
c. It forms the embryo.
d. It protects the embryo.

Solution:

[link] B

Review Questions

Exercise:

Problem: After double fertilization, a zygote and ________ form.

a. an ovule
b. endosperm
c. a cotyledon
d. a suspensor

Solution:

B

Exercise:

Problem: The fertilized ovule gives rise to the ________.

a. fruit
b. seed
c. endosperm
d. embryo

Solution:

B

Exercise:

Problem:

What is the term for a fruit that develops from tissues other than the ovary?

a. simple fruit
b. aggregate fruit
c. multiple fruit
d. accessory fruit

Solution:

D

Exercise:

Problem: The _________ is the outermost covering of a fruit.

a. endocarp
b. pericarp
c. exocarp
d. mesocarp

Solution:

C

Free Response

Exercise:

Problem:

Why do some seeds undergo a period of dormancy, and how do they break dormancy?

Solution:

Many seeds enter a period of inactivity or extremely low metabolic activity, a process known as dormancy. Dormancy allows seeds to tide over unfavorable conditions and germinate on return to favorable conditions. Favorable conditions could be as diverse as moisture, light, cold, fire, or chemical treatments. After heavy rains, many new seedlings emerge. Forest fires also lead to the emergence of new seedlings.

Exercise:

Problem: Discuss some ways in which fruit seeds are dispersed.

Solution:

Some fruits have built-in mechanisms that allow them to disperse seeds by themselves, but others require the assistance of agents like wind, water, and animals. Fruit that are dispersed by the wind are light in weight and often have wing-like appendages that allow them to be carried by the wind; other have structures resembling a parachute that keep them afloat in the wind. Some fruits, such as those of dandelions, have hairy, weightless structures that allow them to float in the wind. Fruits dispersed by water are light and buoyant, giving them the ability to float; coconuts are one example. Animals and birds eat fruits and disperse their seeds by leaving droppings at distant locations. Other animals bury fruit that may later germinate. Some fruits stick to animals' bodies and are carried to new locations. People also contribute to seed dispersal when they carry fruits to new places.

Glossary

accessory fruit
: fruit derived from tissues other than the ovary

aggregate fruit
: fruit that develops from multiple carpels in the same flower

aleurone
: single layer of cells just inside the seed coat that secretes enzymes upon germination

coleoptile
: covering of the shoot tip, found in germinating monocot seeds

coleorhiza
: covering of the root tip, found in germinating monocot seeds

cotyledon
: fleshy part of seed that provides nutrition to the seed

cross-pollination
: transfer of pollen from the anther of one flower to the stigma of a different flower

dormancy
: period of no growth and very slow metabolic processes

double fertilization
: two fertilization events in angiosperms; one sperm fuses with the egg, forming the zygote, whereas the other sperm fuses with the polar nuclei, forming endosperm

endocarp
: innermost part of fruit

endosperm
: triploid structure resulting from fusion of a sperm with polar nuclei, which serves as a nutritive tissue for embryo

endospermic dicot

dicot that stores food reserves in the endosperm

exocarp
outermost covering of a fruit

epicotyl
embryonic shoot above the cotyledons

gravitropism
response of a plant growth in the same direction as gravity

hypocotyl
embryonic axis above the cotyledons

mesocarp
middle part of a fruit

multiple fruit
fruit that develops from multiple flowers on an inflorescence

nectar guide
pigment pattern on a flower that guides an insect to the nectaries

non-endospermic dicot
dicot that stores food reserves in the developing cotyledon

pericarp
collective term describing the exocarp, mesocarp, and endocarp; the structure that encloses the seed and is a part of the fruit

plumule
shoot that develops from the germinating seed

pollination
transfer of pollen to the stigma

radicle
original root that develops from the germinating seed

scarification
: mechanical or chemical processes to soften the seed coat

scutellum
: type of cotyledon found in monocots, as in grass seeds

self-pollination
: transfer of pollen from the anther to the stigma of same flower

simple fruit
: fruit that develops from a single carpel or fused carpels

suspensor
: part of the growing embryo that makes connection with the maternal tissues

tegmen
: inner layer of the seed coat

testa
: outer layer of the seed coat

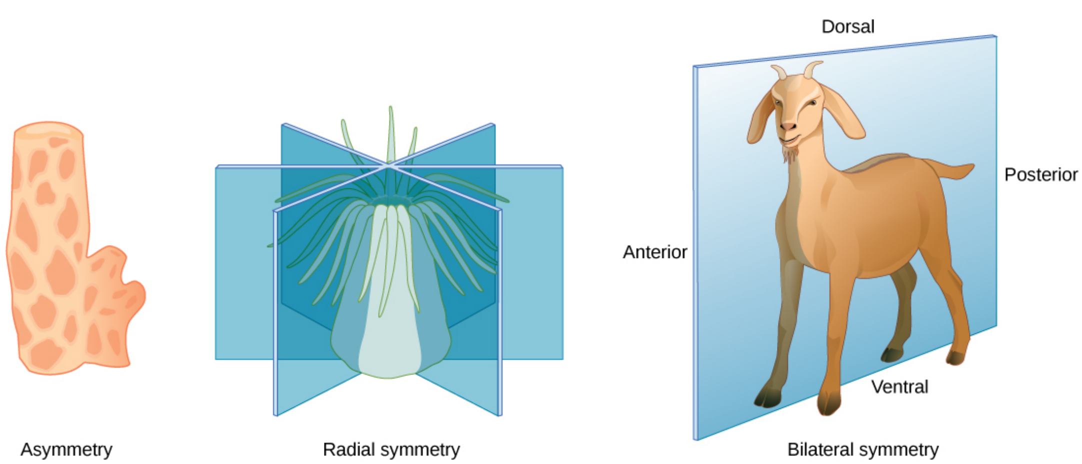
vernalization
: exposure to cold required by some seeds before they can germinate

Animal Form and Function
By the end of this section, you will be able to:

- Describe the various types of body plans that occur in animals
- Describe limits on animal size and shape
- Relate bioenergetics to body size, levels of activity, and the environment

Animals vary in form and function. From a sponge to a worm to a goat, an organism has a distinct body plan that limits its size and shape. Animals' bodies are also designed to interact with their environments, whether in the deep sea, a rainforest canopy, or the desert. Therefore, a large amount of information about the structure of an organism's body (anatomy) and the function of its cells, tissues and organs (physiology) can be learned by studying that organism's environment.

Body Plans



Animals exhibit different types of body symmetry. The sponge is asymmetrical, the sea anemone has radial symmetry, and the goat has bilateral symmetry.

Animal body plans follow set patterns related to symmetry. They are asymmetrical, radial, or bilateral in form as illustrated in [link]. **Asymmetrical** animals are animals with no pattern or symmetry; an example of an asymmetrical animal is a sponge. Radial symmetry, as illustrated in [link], describes when an animal has an up-and-down orientation: any plane cut along its longitudinal axis through the organism produces equal halves, but not a definite right or left side. This plan is found mostly in aquatic animals, especially organisms that attach themselves to a base, like a rock or a boat, and extract their food from the surrounding water as it flows around the organism. Bilateral symmetry is illustrated in the same figure by a goat. The goat also has an upper and lower component to it, but a plane cut from front to back separates the animal into definite right and left sides. Additional terms used when describing positions in the body are anterior (front), posterior (rear), dorsal (toward the back), and ventral (toward the stomach). Bilateral symmetry is found in both land-based and aquatic animals; it enables a high level of mobility.

Limits on Animal Size and Shape

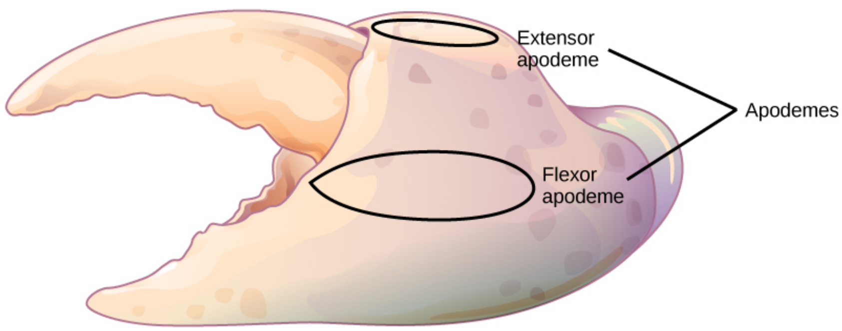
Animals with bilateral symmetry that live in water tend to have a **fusiform** shape: this is a tubular shaped body that is tapered at both ends. This shape decreases the drag on the body as it moves through water and allows the animal to swim at high speeds. [link] lists the maximum speed of various animals. Certain types of sharks can swim at fifty kilometers an hour and some dolphins at 32 to 40 kilometers per hour. Land animals frequently travel faster, although the tortoise and snail are significantly slower than cheetahs. Another difference in the adaptations of aquatic and land-dwelling organisms is that aquatic organisms are constrained in shape by the forces of drag in the water since water has higher viscosity than air. On the other hand, land-dwelling organisms are constrained mainly by gravity, and drag is relatively unimportant. For example, most adaptations in birds are for gravity not for drag.

Maximum Speed of Assorted Land Marine Animals		
Animal	**Speed (kmh)**	**Speed (mph)**
Cheetah	113	70
Quarter horse	77	48
Fox	68	42
Shortfin mako shark	50	31
Domestic house cat	48	30
Human	45	28
Dolphin	32–40	20–25
Mouse	13	8
Snail	0.05	0.03

Most animals have an exoskeleton, including insects, spiders, scorpions, horseshoe crabs, centipedes, and crustaceans. Scientists estimate that, of insects alone, there are over 30 million species on our planet. The exoskeleton is a hard covering or shell that provides benefits to the animal, such as protection against damage from predators and from water loss (for land animals); it also provides for the attachments of muscles.

As the tough and resistant outer cover of an arthropod, the exoskeleton may be constructed of a tough polymer such as chitin and is often biomineralized with materials such as calcium carbonate. This is fused to the animal's epidermis. Ingrowths of the exoskeleton, called **apodemes**, function as attachment sites for muscles, similar to tendons in more advanced animals ([link]). In order to grow, the animal must first synthesize a new exoskeleton underneath the old one and then shed or molt the original

covering. This limits the animal’s ability to grow continually, and may limit the individual’s ability to mature if molting does not occur at the proper time. The thickness of the exoskeleton must be increased significantly to accommodate any increase in weight. It is estimated that a doubling of body size increases body weight by a factor of eight. The increasing thickness of the chitin necessary to support this weight limits most animals with an exoskeleton to a relatively small size. The same principles apply to endoskeletons, but they are more efficient because muscles are attached on the outside, making it easier to compensate for increased mass.



Apodemes are ingrowths on arthropod exoskeletons to which muscles attach. The apodemes on this crab leg are located above and below the fulcrum of the claw. Contraction of muscles attached to the apodemes pulls the claw closed.

An animal with an endoskeleton has its size determined by the amount of skeletal system it needs in order to support the other tissues and the amount of muscle it needs for movement. As the body size increases, both bone and muscle mass increase. The speed achievable by the animal is a balance between its overall size and the bone and muscle that provide support and movement.

Limiting Effects of Diffusion on Size and Development

The exchange of nutrients and wastes between a cell and its watery environment occurs through the process of diffusion. All living cells are bathed in liquid, whether they are in a single-celled organism or a multicellular one. Diffusion is effective over a specific distance and limits the size that an individual cell can attain. If a cell is a single-celled microorganism, such as an amoeba, it can satisfy all of its nutrient and waste needs through diffusion. If the cell is too large, then diffusion is ineffective and the center of the cell does not receive adequate nutrients nor is it able to effectively dispel its waste.

An important concept in understanding how efficient diffusion is as a means of transport is the surface to volume ratio. Recall that any three-dimensional object has a surface area and volume; the ratio of these two quantities is the surface-to-volume ratio. Consider a cell shaped like a perfect sphere: it has a surface area of $4\pi r^2$, and a volume of $(4/3)\pi r^3$. The surface-to-volume ratio of a sphere is $3/r$; as the cell gets bigger, its surface to volume ratio decreases, making diffusion less efficient. The larger the size of the sphere, or animal, the less surface area for diffusion it possesses.

The solution to producing larger organisms is for them to become multicellular. Specialization occurs in complex organisms, allowing cells to become more efficient at doing fewer tasks. For example, circulatory systems bring nutrients and remove waste, while respiratory systems provide oxygen for the cells and remove carbon dioxide from them. Other organ systems have developed further specialization of cells and tissues and efficiently control body functions. Moreover, surface-to-volume ratio applies to other areas of animal development, such as the relationship between muscle mass and cross-sectional surface area in supporting skeletons, and in the relationship between muscle mass and the generation of dissipation of heat.

Note:
Link to Learning

Visit this interactive site to see an entire animal (a zebrafish embryo) at the cellular and sub-cellular level. Use the zoom and navigation functions for a virtual nanoscopy exploration.

Animal Bioenergetics

All animals must obtain their energy from food they ingest or absorb. These nutrients are converted to adenosine triphosphate (ATP) for short-term storage and use by all cells. Some animals store energy for slightly longer times as glycogen, and others store energy for much longer times in the form of triglycerides housed in specialized adipose tissues. No energy system is one hundred percent efficient, and an animal's metabolism produces waste energy in the form of heat. If an animal can conserve that heat and maintain a relatively constant body temperature, it is classified as a warm-blooded animal and called an **endotherm**. The insulation used to conserve the body heat comes in the forms of fur, fat, or feathers. The absence of insulation in **ectothermic** animals increases their dependence on the environment for body heat.

The amount of energy expended by an animal over a specific time is called its metabolic rate. The rate is measured variously in joules, calories, or kilocalories (1000 calories). Carbohydrates and proteins contain about 4.5 to 5 kcal/g, and fat contains about 9 kcal/g. Metabolic rate is estimated as the **basal metabolic rate (BMR)** in endothermic animals at rest and as the **standard metabolic rate (SMR)** in ectotherms. Human males have a BMR of 1600 to 1800 kcal/day, and human females have a BMR of 1300 to 1500 kcal/day. Even with insulation, endothermal animals require extensive amounts of energy to maintain a constant body temperature. An ectotherm such as an alligator has an SMR of 60 kcal/day.

Energy Requirements Related to Body Size

Smaller endothermic animals have a greater surface area for their mass than larger ones ([link]). Therefore, smaller animals lose heat at a faster rate than larger animals and require more energy to maintain a constant internal temperature. This results in a smaller endothermic animal having a higher BMR, per body weight, than a larger endothermic animal.

Species		
Mass	35 g	4,500,000 g
Metabolic rate	890 mm^3 0_2/g body mass/hr	75 mm^3 0_2/g body mass/hr

The mouse has a much higher metabolic rate than the elephant. (credit "mouse": modification of work by Magnus Kjaergaard; credit "elephant": modification of work by "TheLizardQueen"/Flickr)

Energy Requirements Related to Levels of Activity

The more active an animal is, the more energy is needed to maintain that activity, and the higher its BMR or SMR. The average daily rate of energy consumption is about two to four times an animal's BMR or SMR. Humans are more sedentary than most animals and have an average daily rate of only 1.5 times the BMR. The diet of an endothermic animal is determined by its BMR. For example: the type of grasses, leaves, or shrubs that an herbivore eats affects the number of calories that it takes in. The relative caloric content of herbivore foods, in descending order, is tall grasses >

legumes > short grasses > forbs (any broad-leaved plant, not a grass) > subshrubs > annuals/biennials.

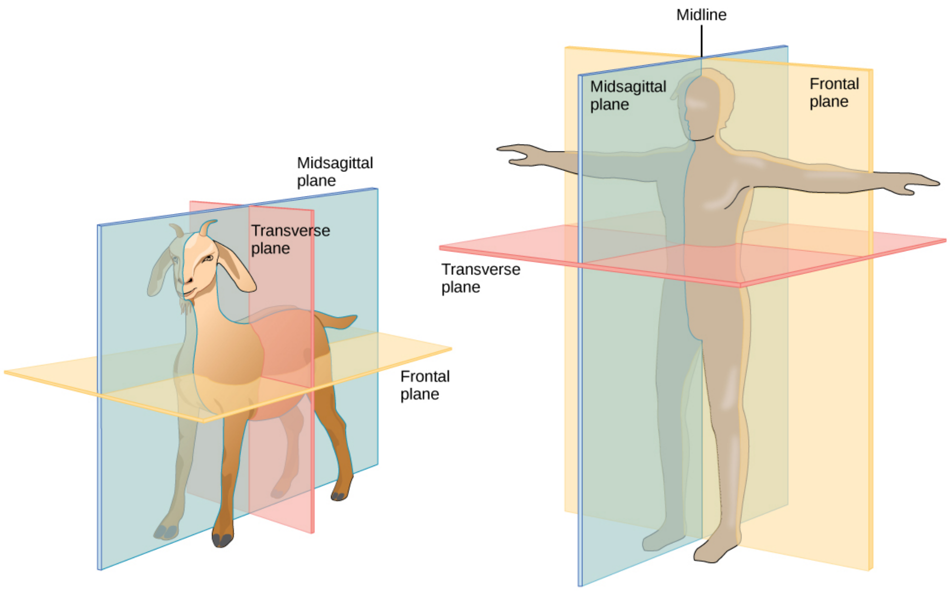
Energy Requirements Related to Environment

Animals adapt to extremes of temperature or food availability through torpor. **Torpor** is a process that leads to a decrease in activity and metabolism and allows animals to survive adverse conditions. Torpor can be used by animals for long periods, such as entering a state of **hibernation** during the winter months, in which case it enables them to maintain a reduced body temperature. During hibernation, ground squirrels can achieve an abdominal temperature of 0° C (32° F), while a bear's internal temperature is maintained higher at about 37° C (99° F).

If torpor occurs during the summer months with high temperatures and little water, it is called **estivation**. Some desert animals use this to survive the harshest months of the year. Torpor can occur on a daily basis; this is seen in bats and hummingbirds. While endothermy is limited in smaller animals by surface to volume ratio, some organisms can be smaller and still be endotherms because they employ daily torpor during the part of the day that is coldest. This allows them to conserve energy during the colder parts of the day, when they consume more energy to maintain their body temperature.

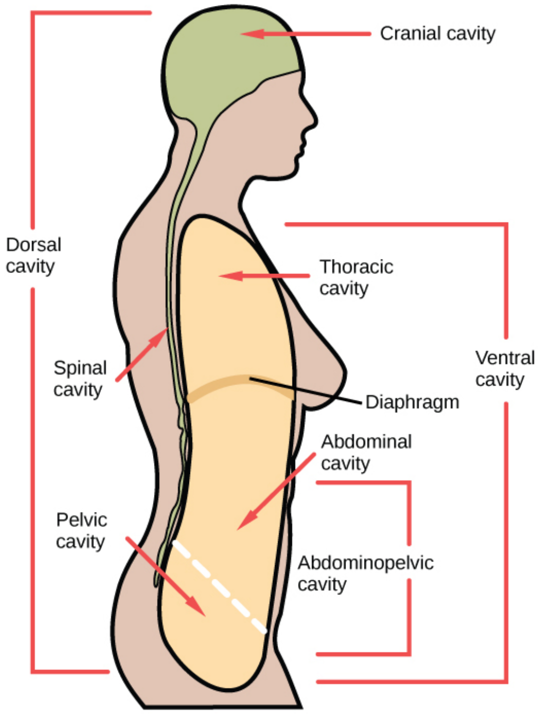
Animal Body Planes and Cavities

A standing vertebrate animal can be divided by several planes. A **sagittal plane** divides the body into right and left portions. A **midsagittal plane** divides the body exactly in the middle, making two equal right and left halves. A **frontal plane** (also called a coronal plane) separates the front from the back. A **transverse plane** (or, horizontal plane) divides the animal into upper and lower portions. This is sometimes called a cross section, and, if the transverse cut is at an angle, it is called an oblique plane. [link] illustrates these planes on a goat (a four-legged animal) and a human being.



Shown are the planes of a quadruped goat and a bipedal human. The midsagittal plane divides the body exactly in half, into right and left portions. The frontal plane divides the front and back, and the transverse plane divides the body into upper and lower portions.

Vertebrate animals have a number of defined body cavities, as illustrated in [link]. Two of these are major cavities that contain smaller cavities within them. The **dorsal cavity** contains the cranial and the vertebral (or spinal) cavities. The **ventral cavity** contains the thoracic cavity, which in turn contains the pleural cavity around the lungs and the pericardial cavity, which surrounds the heart. The ventral cavity also contains the abdominopelvic cavity, which can be separated into the abdominal and the pelvic cavities.



Vertebrate animals have two major body cavities. The dorsal cavity, indicated in green, contains the cranial and the spinal cavity. The ventral cavity, indicated in yellow, contains the thoracic cavity and the abdominopelvic cavity. The thoracic cavity is separated from the abdominopelvic cavity by the diaphragm. The thoracic cavity is separated into the abdominal cavity and the pelvic cavity by an imaginary line parallel to the pelvis bones. (credit: modification of work by NCI)

Note:
Career Connections
Physical Anthropologist
Physical anthropologists study the adaption, variability, and evolution of human beings, plus their living and fossil relatives. They can work in a variety of settings, although most will have an academic appointment at a university, usually in an anthropology department or a biology, genetics, or zoology department.
Non-academic positions are available in the automotive and aerospace industries where the focus is on human size, shape, and anatomy. Research by these professionals might range from studies of how the human body reacts to car crashes to exploring how to make seats more comfortable. Other non-academic positions can be obtained in museums of natural history, anthropology, archaeology, or science and technology. These positions involve educating students from grade school through graduate school. Physical anthropologists serve as education coordinators, collection managers, writers for museum publications, and as administrators. Zoos employ these professionals, especially if they have an expertise in primate biology; they work in collection management and captive breeding programs for endangered species. Forensic science utilizes physical anthropology expertise in identifying human and animal remains, assisting in determining the cause of death, and for expert testimony in trials.

Section Summary

Animal bodies come in a variety of sizes and shapes. Limits on animal size and shape include impacts to their movement. Diffusion affects their size and development. Bioenergetics describes how animals use and obtain energy in relation to their body size, activity level, and environment.

Review Questions

Exercise:

Problem:

Which type of animal maintains a constant internal body temperature?

a. endotherm
b. ectotherm
c. coelomate
d. mesoderm

Solution:

A

Exercise:

Problem:

The symmetry found in animals that move swiftly is _________.

a. radial
b. bilateral
c. sequential
d. interrupted

Solution:

B

Exercise:

Problem:

What term describes the condition of a desert mouse that lowers its metabolic rate and “sleeps” during the hot day?

a. turgid
b. hibernation

c. estivation
d. normal sleep pattern

Solution:

C

Exercise:

Problem:

A plane that divides an animal into equal right and left portions is _________.

a. diagonal
b. midsagittal
c. coronal
d. transverse

Solution:

B

Exercise:

Problem:

A plane that divides an animal into dorsal and ventral portions is _________.

a. sagittal
b. midsagittal
c. coronal
d. transverse

Solution:

D

Exercise:

Problem: The pleural cavity is a part of which cavity?

a. dorsal cavity
b. thoracic cavity
c. abdominal cavity
d. pericardial cavity

Solution:

B

Free Response

Exercise:

Problem:

How does diffusion limit the size of an organism? How is this counteracted?

Solution:

Diffusion is effective over a very short distance. If a cell exceeds this distance in its size, the center of the cell cannot get adequate nutrients nor can it expel enough waste to survive. To compensate for this, cells can loosely adhere to each other in a liquid medium, or develop into multi-celled organisms that use circulatory and respiratory systems to deliver nutrients and remove wastes.

Exercise:

Problem: What is the relationship between BMR and body size? Why?

Solution:

Basal Metabolic Rate is an expression of the metabolic processes that occur to maintain an individual’s functioning and body temperature. Smaller bodied animals have a relatively large surface area compared to a much larger animal. The small animal’s large surface area leads to increased heat loss that the animal must compensate for, resulting in a higher BMR. A large animal, having less relative surface area, does not lose as much heat and has a correspondingly lower BMR.

Glossary

apodeme
: ingrowth of an animal’s exoskeleton that functions as an attachment site for muscles

asymmetrical
: describes animals with no axis of symmetry in their body pattern

basal metabolic rate (BMR)
: metabolic rate at rest in endothermic animals

dorsal cavity
: body cavity on the posterior or back portion of an animal; includes the cranial and vertebral cavities

ectotherm
: animal incapable of maintaining a relatively constant internal body temperature

endotherm
: animal capable of maintaining a relatively constant internal body temperature

estivation
: torpor in response to extremely high temperatures and low water availability

frontal (coronal) plane

plane cutting through an animal separating the individual into front and back portions

fusiform
: animal body shape that is tubular and tapered at both ends

hibernation
: torpor over a long period of time, such as a winter

midsagittal plane
: plane cutting through an animal separating the individual into even right and left sides

sagittal plane
: plane cutting through an animal separating the individual into right and left sides

standard metabolic rate (SMR)
: metabolic rate at rest in ectothermic animals

torpor
: decrease in activity and metabolism that allows an animal to survive adverse conditions

transverse (horizontal) plane
: plane cutting through an animal separating the individual into upper and lower portions

ventral cavity
: body cavity on the anterior or front portion of an animal that includes the thoracic cavities and the abdominopelvic cavities

Animal Primary Tissues
By the end of this section, you will be able to:

- Describe epithelial tissues
- Discuss the different types of connective tissues in animals
- Describe three types of muscle tissues
- Describe nervous tissue

The tissues of multicellular, complex animals are four primary types: epithelial, connective, muscle, and nervous. Recall that tissues are groups of similar cells group of similar cells carrying out related functions. These tissues combine to form organs—like the skin or kidney—that have specific, specialized functions within the body. Organs are organized into organ systems to perform functions; examples include the circulatory system, which consists of the heart and blood vessels, and the digestive system, consisting of several organs, including the stomach, intestines, liver, and pancreas. Organ systems come together to create an entire organism.

Epithelial Tissues

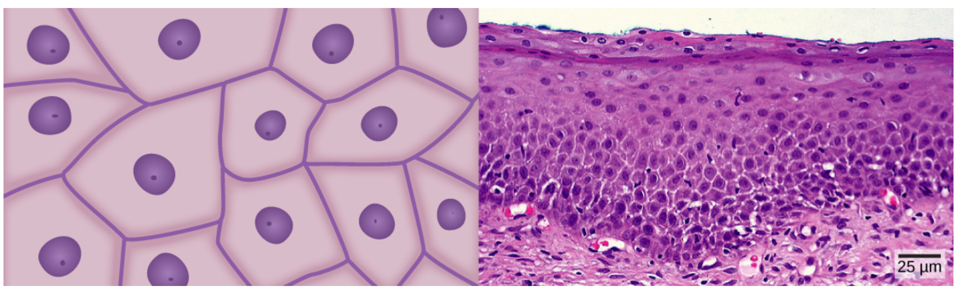
Epithelial tissues cover the outside of organs and structures in the body and line the lumens of organs in a single layer or multiple layers of cells. The types of epithelia are classified by the shapes of cells present and the number of layers of cells. Epithelia composed of a single layer of cells is called **simple epithelia**; epithelial tissue composed of multiple layers is called **stratified epithelia**. [link] summarizes the different types of epithelial tissues.

Different Types of Epithelial Tissues		
Cell shape	**Description**	**Location**

Different Types of Epithelial Tissues		
Cell shape	**Description**	**Location**
squamous	flat, irregular round shape	simple: lung alveoli, capillaries stratified: skin, mouth, vagina
cuboidal	cube shaped, central nucleus	glands, renal tubules
columnar	tall, narrow, nucleus toward base tall, narrow, nucleus along cell	simple: digestive tract pseudostratified: respiratory tract
transitional	round, simple but appear stratified	urinary bladder

Squamous Epithelia

Squamous epithelial cells are generally round, flat, and have a small, centrally located nucleus. The cell outline is slightly irregular, and cells fit together to form a covering or lining. When the cells are arranged in a single layer (simple epithelia), they facilitate diffusion in tissues, such as the areas of gas exchange in the lungs and the exchange of nutrients and waste at blood capillaries.

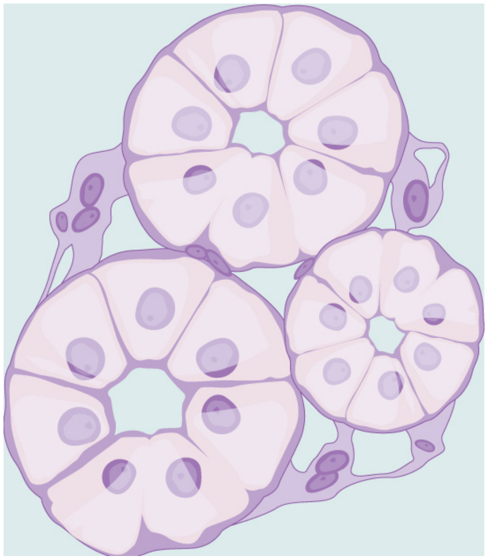


Squamous epithelia cells (a) have a slightly irregular shape, and a small, centrally located nucleus. These cells can be stratified into layers, as in (b) this human cervix specimen. (credit b: modification of work by Ed Uthman; scale-bar data from Matt Russell)

[link]**a** illustrates a layer of squamous cells with their membranes joined together to form an epithelium. Image [link]**b** illustrates squamous epithelial cells arranged in stratified layers, where protection is needed on the body from outside abrasion and damage. This is called a stratified squamous epithelium and occurs in the skin and in tissues lining the mouth and vagina.

Cuboidal Epithelia

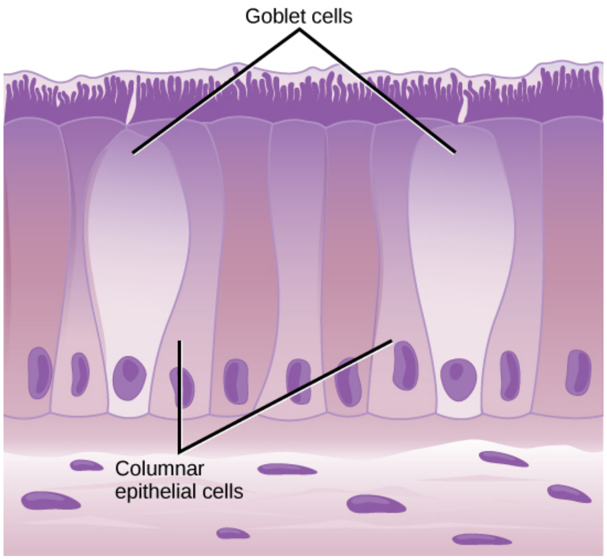
Cuboidal epithelial cells, shown in [link], are cube-shaped with a single, central nucleus. They are most commonly found in a single layer representing a simple epithelia in glandular tissues throughout the body where they prepare and secrete glandular material. They are also found in the walls of tubules and in the ducts of the kidney and liver.



Simple cuboidal epithelial cells line tubules in the mammalian kidney, where they are involved in filtering the blood.

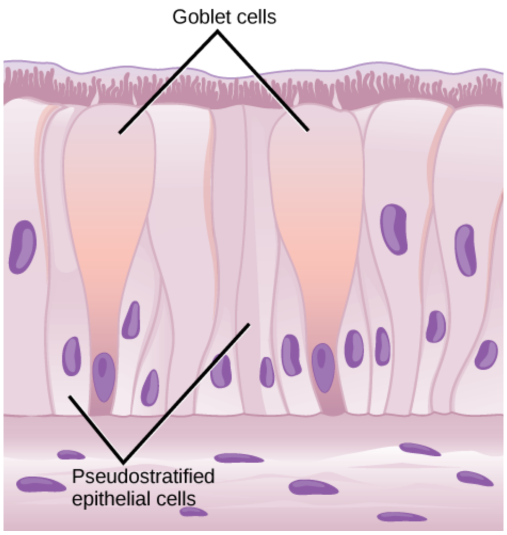
Columnar Epithelia

Columnar epithelial cells are taller than they are wide: they resemble a stack of columns in an epithelial layer, and are most commonly found in a single-layer arrangement. The nuclei of columnar epithelial cells in the digestive tract appear to be lined up at the base of the cells, as illustrated in [link]. These cells absorb material from the lumen of the digestive tract and prepare it for entry into the body through the circulatory and lymphatic systems.



Simple columnar epithelial cells absorb material from the digestive tract. Goblet cells secret mucous into the digestive tract lumen.

Columnar epithelial cells lining the respiratory tract appear to be stratified. However, each cell is attached to the base membrane of the tissue and, therefore, they are simple tissues. The nuclei are arranged at different levels in the layer of cells, making it appear as though there is more than one layer, as seen in [link]. This is called **pseudostratified**, columnar epithelia. This cellular covering has cilia at the apical, or free, surface of the cells. The cilia enhance the movement of mucous and trapped particles out of the respiratory tract, helping to protect the system from invasive microorganisms and harmful material that has been breathed into the body. Goblet cells are interspersed in some tissues (such as the lining of the trachea). The goblet cells contain mucous that traps irritants, which in the case of the trachea keep these irritants from getting into the lungs.



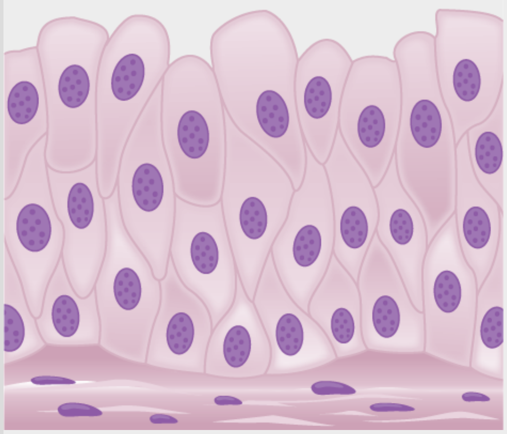
Pseudostratified columnar epithelia line the respiratory tract. They exist in one layer, but the arrangement of nuclei at different levels makes it appear that there is more than one layer. Goblet cells interspersed between the columnar epithelial cells secrete mucous into the respiratory tract.

Transitional Epithelia

Transitional or uroepithelial cells appear only in the urinary system, primarily in the bladder and ureter. These cells are arranged in a stratified layer, but they have the capability of appearing to pile up on top of each other in a relaxed, empty bladder, as illustrated in [link]. As the urinary bladder fills, the epithelial layer unfolds and expands to hold the volume of

urine introduced into it. As the bladder fills, it expands and the lining becomes thinner. In other words, the tissue transitions from thick to thin.

Note:
Art Connection



Transitional epithelia of the urinary bladder undergo changes in thickness depending on how full the bladder is.

Which of the following statements about types of epithelial cells is false?

a. Simple columnar epithelial cells line the tissue of the lung.
b. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
c. Pseudostratisfied columnar epithilia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
d. Transitional epithelia change in thickness depending on how full the bladder is.

Connective Tissues

Connective tissues are made up of a matrix consisting of living cells and a non-living substance, called the ground substance. The ground substance is made of an organic substance (usually a protein) and an inorganic substance (usually a mineral or water). The principal cell of connective tissues is the fibroblast. This cell makes the fibers found in nearly all of the connective tissues. Fibroblasts are motile, able to carry out mitosis, and can synthesize whichever connective tissue is needed. Macrophages, lymphocytes, and, occasionally, leukocytes can be found in some of the tissues. Some tissues have specialized cells that are not found in the others. The **matrix** in connective tissues gives the tissue its density. When a connective tissue has a high concentration of cells or fibers, it has proportionally a less dense matrix.

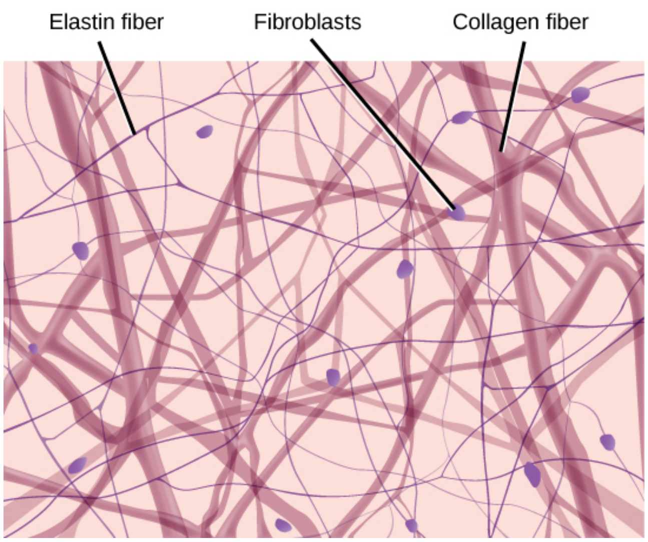
The organic portion or protein fibers found in connective tissues are either collagen, elastic, or reticular fibers. Collagen fibers provide strength to the tissue, preventing it from being torn or separated from the surrounding tissues. Elastic fibers are made of the protein elastin; this fiber can stretch to one and one half of its length and return to its original size and shape. Elastic fibers provide flexibility to the tissues. Reticular fibers are the third type of protein fiber found in connective tissues. This fiber consists of thin strands of collagen that form a network of fibers to support the tissue and other organs to which it is connected. The various types of connective tissues, the types of cells and fibers they are made of, and sample locations of the tissues is summarized in [link].

Connective Tissues			
Tissue	**Cells**	**Fibers**	**Location**

Connective Tissues			
Tissue	**Cells**	**Fibers**	**Location**
loose/areolar	fibroblasts, macrophages, some lymphocytes, some neutrophils	few: collagen, elastic, reticular	around blood vessels; anchors epithelia
dense, fibrous connective tissue	fibroblasts, macrophages,	mostly collagen	irregular: skin regular: tendons, ligaments
cartilage	chondrocytes, chondroblasts	hyaline: few collagen fibrocartilage: large amount of collagen	shark skeleton, fetal bones, human ears, intervertebral discs
bone	osteoblasts, osteocytes, osteoclasts	some: collagen, elastic	vertebrate skeletons
adipose	adipocytes	few	adipose (fat)
blood	red blood cells, white blood cells	none	blood

Loose/Areolar Connective Tissue

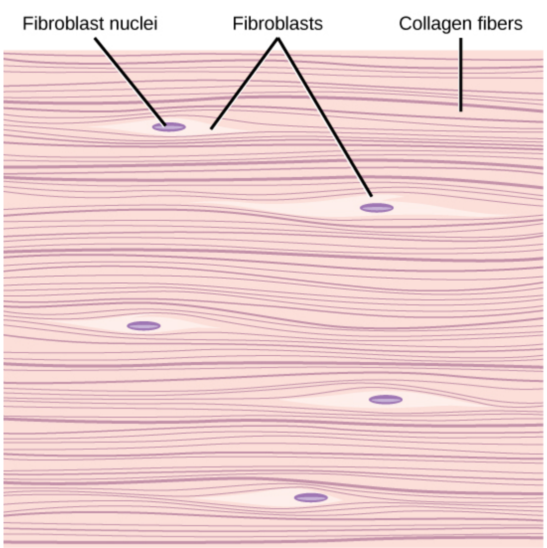
Loose connective tissue, also called areolar connective tissue, has a sampling of all of the components of a connective tissue. As illustrated in [link], loose connective tissue has some fibroblasts; macrophages are present as well. Collagen fibers are relatively wide and stain a light pink, while elastic fibers are thin and stain dark blue to black. The space between the formed elements of the tissue is filled with the matrix. The material in the connective tissue gives it a loose consistency similar to a cotton ball that has been pulled apart. Loose connective tissue is found around every blood vessel and helps to keep the vessel in place. The tissue is also found around and between most body organs. In summary, areolar tissue is tough, yet flexible, and comprises membranes.



Loose connective tissue is composed of loosely woven collagen and elastic fibers. The fibers and other components of the connective tissue matrix are secreted by fibroblasts.

Fibrous Connective Tissue

Fibrous connective tissues contain large amounts of collagen fibers and few cells or matrix material. The fibers can be arranged irregularly or regularly with the strands lined up in parallel. Irregularly arranged fibrous connective tissues are found in areas of the body where stress occurs from all directions, such as the dermis of the skin. Regular fibrous connective tissue, shown in [link], is found in tendons (which connect muscles to bones) and ligaments (which connect bones to bones).



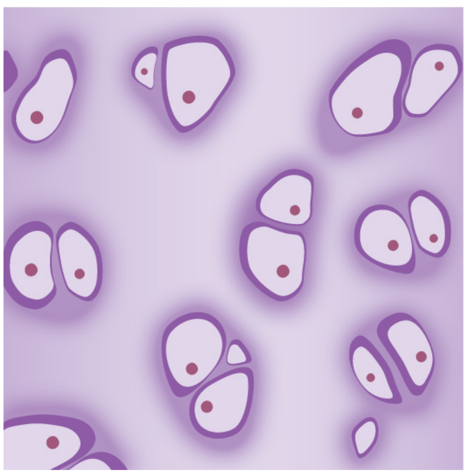
Fibrous connective tissue from the tendon has strands of collagen fibers lined up in parallel.

Cartilage

Cartilage is a connective tissue with a large amount of the matrix and variable amounts of fibers. The cells, called **chondrocytes**, make the matrix

and fibers of the tissue. Chondrocytes are found in spaces within the tissue called **lacunae**.

A cartilage with few collagen and elastic fibers is hyaline cartilage, illustrated in [link]. The lacunae are randomly scattered throughout the tissue and the matrix takes on a milky or scrubbed appearance with routine histological stains. Sharks have cartilaginous skeletons, as does nearly the entire human skeleton during a specific pre-birth developmental stage. A remnant of this cartilage persists in the outer portion of the human nose. Hyaline cartilage is also found at the ends of long bones, reducing friction and cushioning the articulations of these bones.



Hyaline cartilage consists of a matrix with cells called chondrocytes embedded in it. The chondrocytes exist in cavities in the matrix called lacunae.

Elastic cartilage has a large amount of elastic fibers, giving it tremendous flexibility. The ears of most vertebrate animals contain this cartilage as do portions of the larynx, or voice box. Fibrocartilage contains a large amount of collagen fibers, giving the tissue tremendous strength. Fibrocartilage comprises the intervertebral discs in vertebrate animals. Hyaline cartilage found in movable joints such as the knee and shoulder becomes damaged as a result of age or trauma. Damaged hyaline cartilage is replaced by fibrocartilage and results in the joints becoming “stiff.”

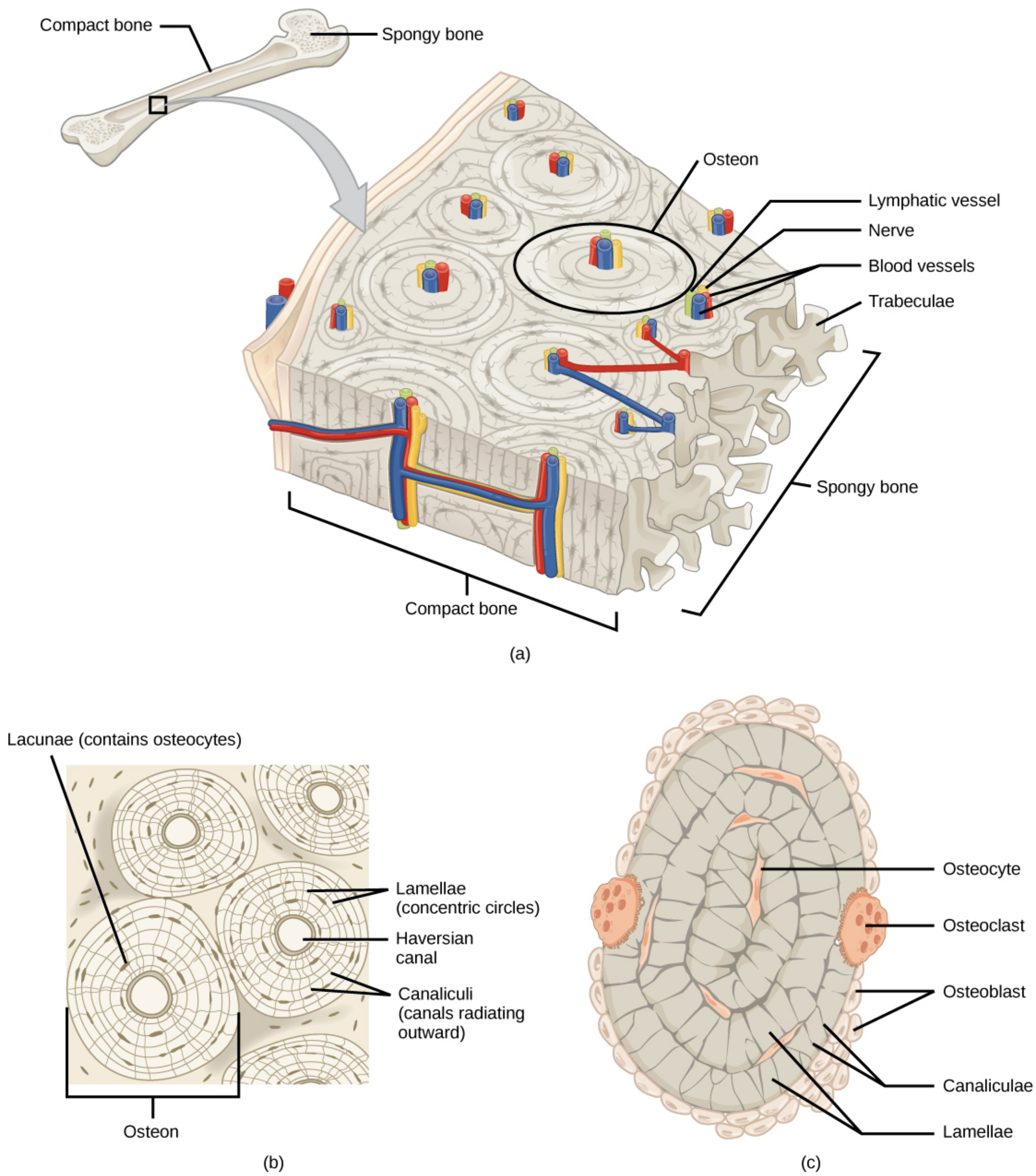
Bone

Bone, or osseous tissue, is a connective tissue that has a large amount of two different types of matrix material. The organic matrix is similar to the matrix material found in other connective tissues, including some amount of collagen and elastic fibers. This gives strength and flexibility to the tissue. The inorganic matrix consists of mineral salts—mostly calcium salts—that give the tissue hardness. Without adequate organic material in the matrix, the tissue breaks; without adequate inorganic material in the matrix, the tissue bends.

There are three types of cells in bone: osteoblasts, osteocytes, and osteoclasts. Osteoblasts are active in making bone for growth and remodeling. Osteoblasts deposit bone material into the matrix and, after the matrix surrounds them, they continue to live, but in a reduced metabolic state as osteocytes. Osteocytes are found in lacunae of the bone. Osteoclasts are active in breaking down bone for bone remodeling, and they provide access to calcium stored in tissues. Osteoclasts are usually found on the surface of the tissue.

Bone can be divided into two types: compact and spongy. Compact bone is found in the shaft (or diaphysis) of a long bone and the surface of the flat bones, while spongy bone is found in the end (or epiphysis) of a long bone. Compact bone is organized into subunits called **osteons**, as illustrated in [link]. A blood vessel and a nerve are found in the center of the structure within the Haversian canal, with radiating circles of lacunae around it known as lamellae. The wavy lines seen between the lacunae are

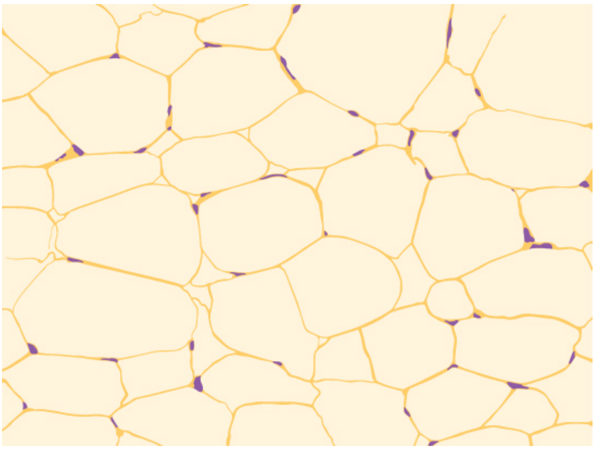
microchannels called **canaliculi**; they connect the lacunae to aid diffusion between the cells. Spongy bone is made of tiny plates called **trabeculae** these plates serve as struts to give the spongy bone strength. Over time, these plates can break causing the bone to become less resilient. Bone tissue forms the internal skeleton of vertebrate animals, providing structure to the animal and points of attachment for tendons.



(a) Compact bone is a dense matrix on the outer surface of bone. Spongy bone, inside the compact bone, is porous with web-like trabeculae. (b) Compact bone is organized into rings called osteons. Blood vessels, nerves, and lymphatic vessels are found in the central Haversian canal. Rings of lamellae surround the Haversian canal. Between the lamellae are cavities called lacunae. Canaliculi are microchannels connecting the lacunae together. (c) Osteoblasts surround the exterior of the bone. Osteoclasts bore tunnels into the bone and osteocytes are found in the lacunae.

Adipose Tissue

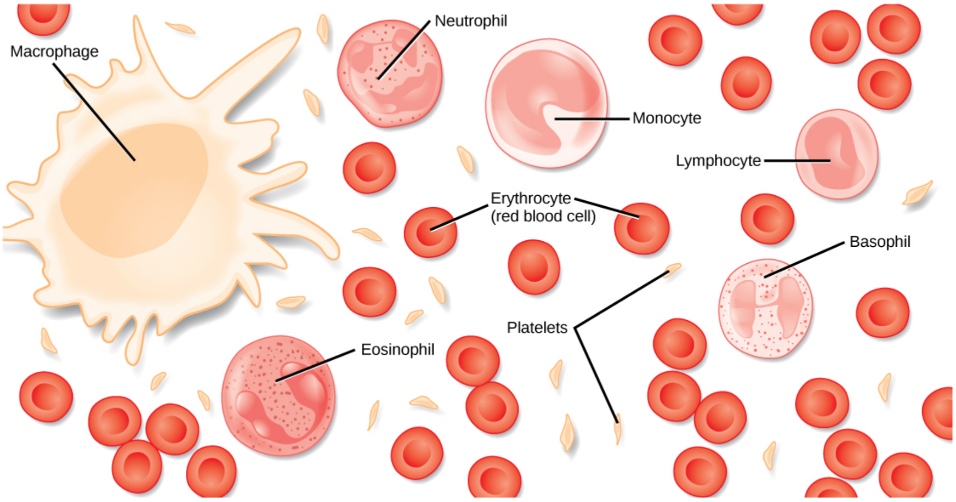
Adipose tissue, or fat tissue, is considered a connective tissue even though it does not have fibroblasts or a real matrix and only has a few fibers. Adipose tissue is made up of cells called adipocytes that collect and store fat in the form of triglycerides, for energy metabolism. Adipose tissues additionally serve as insulation to help maintain body temperatures, allowing animals to be endothermic, and they function as cushioning against damage to body organs. Under a microscope, adipose tissue cells appear empty due to the extraction of fat during the processing of the material for viewing, as seen in [link]. The thin lines in the image are the cell membranes, and the nuclei are the small, black dots at the edges of the cells.



Adipose is a connective tissue is made up of cells called adipocytes. Adipocytes have small nuclei localized at the cell edge.

Blood

Blood is considered a connective tissue because it has a matrix, as shown in [link]. The living cell types are red blood cells (RBC), also called erythrocytes, and white blood cells (WBC), also called leukocytes. The fluid portion of whole blood, its matrix, is commonly called plasma.



Blood is a connective tissue that has a fluid matrix, called plasma, and no fibers. Erythrocytes (red blood cells), the predominant cell type, are involved in the transport of oxygen and carbon dioxide. Also present are various leukocytes (white blood cells) involved in immune response.

The cell found in greatest abundance in blood is the erythrocyte. Erythrocytes are counted in millions in a blood sample: the average number of red blood cells in primates is 4.7 to 5.5 million cells per microliter. Erythrocytes are consistently the same size in a species, but vary in size between species. For example, the average diameter of a primate red blood cell is 7.5 µl, a dog is close at 7.0 µl, but a cat's RBC diameter is 5.9 µl. Sheep erythrocytes are even smaller at 4.6 µl. Mammalian erythrocytes lose their nuclei and mitochondria when they are released from the bone marrow where they are made. Fish, amphibian, and avian red blood cells maintain their nuclei and mitochondria throughout the cell's life. The principal job of an erythrocyte is to carry and deliver oxygen to the tissues.

Leukocytes are the predominant white blood cells found in the peripheral blood. Leukocytes are counted in the thousands in the blood with measurements expressed as ranges: primate counts range from 4,800 to 10,800 cells per µl, dogs from 5,600 to 19,200 cells per µl, cats from 8,000

to 25,000 cells per µl, cattle from 4,000 to 12,000 cells per µl, and pigs from 11,000 to 22,000 cells per µl.

Lymphocytes function primarily in the immune response to foreign antigens or material. Different types of lymphocytes make antibodies tailored to the foreign antigens and control the production of those antibodies. Neutrophils are phagocytic cells and they participate in one of the early lines of defense against microbial invaders, aiding in the removal of bacteria that has entered the body. Another leukocyte that is found in the peripheral blood is the monocyte. Monocytes give rise to phagocytic macrophages that clean up dead and damaged cells in the body, whether they are foreign or from the host animal. Two additional leukocytes in the blood are eosinophils and basophils—both help to facilitate the inflammatory response.

The slightly granular material among the cells is a cytoplasmic fragment of a cell in the bone marrow. This is called a platelet or thrombocyte. Platelets participate in the stages leading up to coagulation of the blood to stop bleeding through damaged blood vessels. Blood has a number of functions, but primarily it transports material through the body to bring nutrients to cells and remove waste material from them.

Muscle Tissues

There are three types of muscle in animal bodies: smooth, skeletal, and cardiac. They differ by the presence or absence of striations or bands, the number and location of nuclei, whether they are voluntarily or involuntarily controlled, and their location within the body. [link] summarizes these differences.

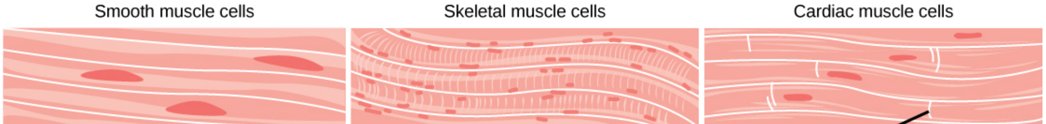
Types of Muscles

Types of Muscles

Type of Muscle	Striations	Nuclei	Control	Location
Type of Muscle	**Striations**	**Nuclei**	**Control**	**Location**
smooth	no	single, in center	involuntary	visceral organs
skeletal	yes	many, at periphery	voluntary	skeletal muscles
cardiac	yes	single, in center	involuntary	heart

Smooth Muscle

Smooth muscle does not have striations in its cells. It has a single, centrally located nucleus, as shown in [link]. Constriction of smooth muscle occurs under involuntary, autonomic nervous control and in response to local conditions in the tissues. Smooth muscle tissue is also called non-striated as it lacks the banded appearance of skeletal and cardiac muscle. The walls of blood vessels, the tubes of the digestive system, and the tubes of the reproductive systems are composed of mostly smooth muscle.



Smooth muscle cells do not have striations, while skeletal

muscle cells do. Cardiac muscle cells have striations, but, unlike the multinucleate skeletal cells, they have only one nucleus. Cardiac muscle tissue also has intercalated discs, specialized regions running along the plasma membrane that join adjacent cardiac muscle cells and assist in passing an electrical impulse from cell to cell.

Skeletal Muscle

Skeletal muscle has striations across its cells caused by the arrangement of the contractile proteins actin and myosin. These muscle cells are relatively long and have multiple nuclei along the edge of the cell. Skeletal muscle is under voluntary, somatic nervous system control and is found in the muscles that move bones. [link] illustrates the histology of skeletal muscle.

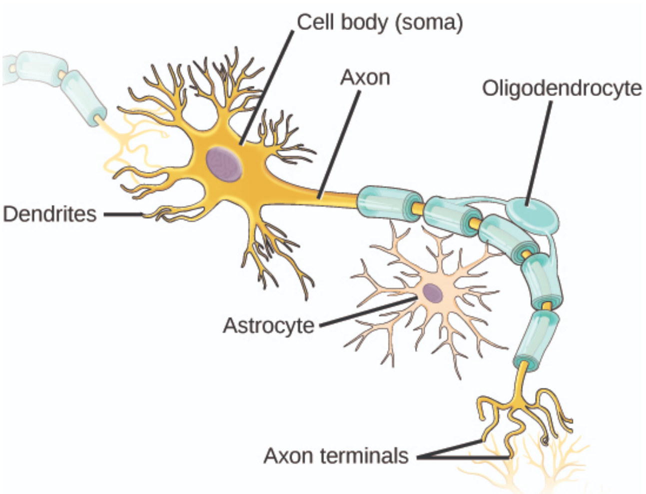
Cardiac Muscle

Cardiac muscle, shown in [link], is found only in the heart. Like skeletal muscle, it has cross striations in its cells, but cardiac muscle has a single, centrally located nucleus. Cardiac muscle is not under voluntary control but can be influenced by the autonomic nervous system to speed up or slow down. An added feature to cardiac muscle cells is a line than extends along the end of the cell as it abuts the next cardiac cell in the row. This line is called an intercalated disc: it assists in passing electrical impulse efficiently from one cell to the next and maintains the strong connection between neighboring cardiac cells.

Nervous Tissues

Nervous tissues are made of cells specialized to receive and transmit electrical impulses from specific areas of the body and to send them to specific locations in the body. The main cell of the nervous system is the

neuron, illustrated in [link]. The large structure with a central nucleus is the cell body of the neuron. Projections from the cell body are either dendrites specialized in receiving input or a single axon specialized in transmitting impulses. Some glial cells are also shown. Astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently. Other glial cells that are not shown support the nutritional and waste requirements of the neuron. Some of the glial cells are phagocytic and remove debris or damaged cells from the tissue. A nerve consists of neurons and glial cells.



The neuron has projections called dendrites that receive signals and projections called axons that send signals. Also shown are two types of glial cells: astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently.

Note:
Link to Learning

Click through the interactive review to learn more about epithelial tissues.

Note:
Career Connections
Pathologist
A pathologist is a medical doctor or veterinarian who has specialized in the laboratory detection of disease in animals, including humans. These professionals complete medical school education and follow it with an extensive post-graduate residency at a medical center. A pathologist may oversee clinical laboratories for the evaluation of body tissue and blood samples for the detection of disease or infection. They examine tissue specimens through a microscope to identify cancers and other diseases. Some pathologists perform autopsies to determine the cause of death and the progression of disease.

Section Summary

The basic building blocks of complex animals are four primary tissues. These are combined to form organs, which have a specific, specialized function within the body, such as the skin or kidney. Organs are organized together to perform common functions in the form of systems. The four primary tissues are epithelia, connective tissues, muscle tissues, and nervous tissues.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about types of epithelial cells is false?

a. Simple columnar epithelial cells line the tissue of the lung.
b. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
c. Pseudostratisfied columnar epithilia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
d. Transitional epithelia change in thickness depending on how full the bladder is.

Solution:

[link] A

Review Questions

Exercise:

Problem: Which type of epithelial cell is best adapted to aid diffusion?

a. squamous
b. cuboidal
c. columnar
d. transitional

Solution:

A

Exercise:

Problem: Which type of epithelial cell is found in glands?

a. squamous
b. cuboidal
c. columnar
d. transitional

Solution:

B

Exercise:

Problem: Which type of epithelial cell is found in the urinary bladder?

a. squamous
b. cuboidal
c. columnar
d. transitional

Solution:

D

Exercise:

Problem: Which type of connective tissue has the most fibers?

a. loose connective tissue
b. fibrous connective tissue
c. cartilage
d. bone

Solution:

B

Exercise:

Problem:

Which type of connective tissue has a mineralized different matrix?

a. loose connective tissue
b. fibrous connective tissue
c. cartilage
d. bone

Solution:

D

Exercise:

Problem:

The cell found in bone that breaks it down is called an _________.

a. osteoblast
b. osteocyte
c. osteoclast
d. osteon

Solution:

C

Exercise:

Problem:

The cell found in bone that makes the bone is called an _________.

a. osteoblast
b. osteocyte
c. osteoclast
d. osteon

Solution:

A

Exercise:

Problem: Plasma is the _________.

a. fibers in blood
b. matrix of blood
c. cell that phagocytizes bacteria
d. cell fragment found in the tissue

Solution:

B

Exercise:

Problem:

The type of muscle cell under voluntary control is the _________.

a. smooth muscle
b. skeletal muscle
c. cardiac muscle
d. visceral muscle

Solution:

B

Exercise:

Problem: The part of a neuron that contains the nucleus is the

a. cell body
b. dendrite
c. axon
d. glial

Solution:

A

Free Response

Exercise:

Problem:

How can squamous epithelia both facilitate diffusion and prevent damage from abrasion?

Solution:

Squamous epithelia can be either simple or stratified. As a single layer of cells, it presents a very thin epithelia that minimally inhibits diffusion. As a stratified epithelia, the surface cells can be sloughed off and the cells in deeper layers protect the underlying tissues from damage.

Exercise:

Problem: What are the similarities between cartilage and bone?

Solution:

Both contain cells other than the traditional fibroblast. Both have cells that lodge in spaces within the tissue called lacunae. Both collagen and elastic fibers are found in bone and cartilage. Both tissues participate in vertebrate skeletal development and formation.

Glossary

canaliculus
: microchannel that connects the lacunae and aids diffusion between cells

cartilage
: type of connective tissue with a large amount of ground substance matrix, cells called chondrocytes, and some amount of fibers

chondrocyte
: cell found in cartilage

columnar epithelia
: epithelia made of cells taller than they are wide, specialized in absorption

connective tissue
: type of tissue made of cells, ground substance matrix, and fibers

cuboidal epithelia
: epithelia made of cube-shaped cells, specialized in glandular functions

epithelial tissue
: tissue that either lines or covers organs or other tissues

fibrous connective tissue
: type of connective tissue with a high concentration of fibers

lacuna
: space in cartilage and bone that contains living cells

loose (areolar) connective tissue
    type of connective tissue with small amounts of cells, matrix, and fibers; found around blood vessels

matrix
    component of connective tissue made of both living and non-living (ground substances) cells

osteon
    subunit of compact bone

pseudostratified
    layer of epithelia that appears multilayered, but is a simple covering

simple epithelia
    single layer of epithelial cells

squamous epithelia
    type of epithelia made of flat cells, specialized in aiding diffusion or preventing abrasion

stratified epithelia
    multiple layers of epithelial cells

trabecula
    tiny plate that makes up spongy bone and gives it strength

transitional epithelia
    epithelia that can transition for appearing multilayered to simple; also called uroepithelial

Homeostasis
By the end of this section, you will be able to:

- Define homeostasis
- Describe the factors affecting homeostasis
- Discuss positive and negative feedback mechanisms used in homeostasis
- Describe thermoregulation of endothermic and ectothermic animals

Animal organs and organ systems constantly adjust to internal and external changes through a process called homeostasis (“steady state”). These changes might be in the level of glucose or calcium in blood or in external temperatures. **Homeostasis** means to maintain dynamic equilibrium in the body. It is dynamic because it is constantly adjusting to the changes that the body’s systems encounter. It is equilibrium because body functions are kept within specific ranges. Even an animal that is apparently inactive is maintaining this homeostatic equilibrium.

Homeostatic Process

The goal of homeostasis is the maintenance of equilibrium around a point or value called a **set point**. While there are normal fluctuations from the set point, the body’s systems will usually attempt to go back to this point. A change in the internal or external environment is called a stimulus and is detected by a receptor; the response of the system is to adjust the deviation parameter toward the set point. For instance, if the body becomes too warm, adjustments are made to cool the animal. If the blood’s glucose rises after a meal, adjustments are made to lower the blood glucose level by getting the nutrient into tissues that need it or to store it for later use.

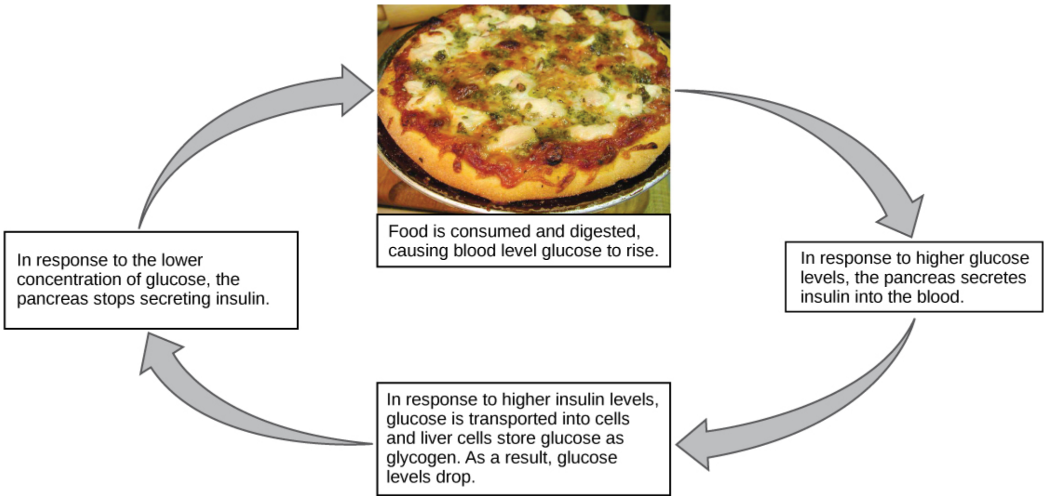
Control of Homeostasis

When a change occurs in an animal’s environment, an adjustment must be made. The receptor senses the change in the environment, then sends a signal to the control center (in most cases, the brain) which in turn generates a response that is signaled to an effector. The effector is a muscle (that contracts or relaxes) or a gland that secretes. Homeostatsis is

maintained by negative feedback loops. Positive feedback loops actually push the organism further out of homeostasis, but may be necessary for life to occur. Homeostasis is controlled by the nervous and endocrine system of mammals.

Negative Feedback Mechanisms

Any homeostatic process that changes the direction of the stimulus is a **negative feedback loop**. It may either increase or decrease the stimulus, but the stimulus is not allowed to continue as it did before the receptor sensed it. In other words, if a level is too high, the body does something to bring it down, and conversely, if a level is too low, the body does something to make it go up. Hence the term negative feedback. An example is animal maintenance of blood glucose levels. When an animal has eaten, blood glucose levels rise. This is sensed by the nervous system. Specialized cells in the pancreas sense this, and the hormone insulin is released by the endocrine system. Insulin causes blood glucose levels to decrease, as would be expected in a negative feedback system, as illustrated in [link]. However, if an animal has not eaten and blood glucose levels decrease, this is sensed in another group of cells in the pancreas, and the hormone glucagon is released causing glucose levels to increase. This is still a negative feedback loop, but not in the direction expected by the use of the term “negative.” Another example of an increase as a result of the feedback loop is the control of blood calcium. If calcium levels decrease, specialized cells in the parathyroid gland sense this and release parathyroid hormone (PTH), causing an increased absorption of calcium through the intestines and kidneys and, possibly, the breakdown of bone in order to liberate calcium. The effects of PTH are to raise blood levels of the element. Negative feedback loops are the predominant mechanism used in homeostasis.

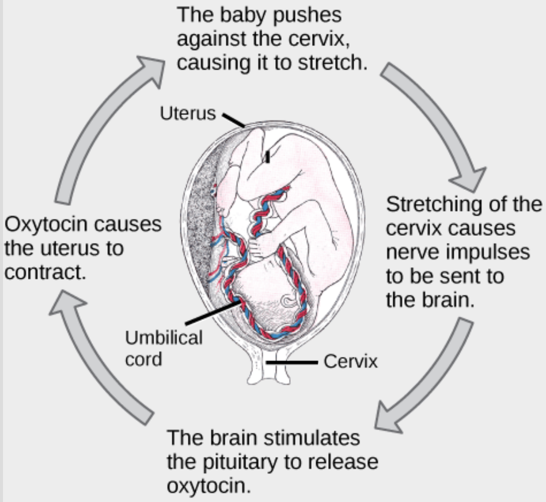


Blood sugar levels are controlled by a negative feedback loop. (credit: modification of work by Jon Sullivan)

Positive Feedback Loop

A **positive feedback loop** maintains the direction of the stimulus, possibly accelerating it. Few examples of positive feedback loops exist in animal bodies, but one is found in the cascade of chemical reactions that result in blood clotting, or coagulation. As one clotting factor is activated, it activates the next factor in sequence until a fibrin clot is achieved. The direction is maintained, not changed, so this is positive feedback. Another example of positive feedback is uterine contractions during childbirth, as illustrated in [link]. The hormone oxytocin, made by the endocrine system, stimulates the contraction of the uterus. This produces pain sensed by the nervous system. Instead of lowering the oxytocin and causing the pain to subside, more oxytocin is produced until the contractions are powerful enough to produce childbirth.

Note:
Art Connection



The birth of a human infant is the result of positive feedback.

State whether each of the following processes is regulated by a positive feedback loop or a negative feedback loop.

a. A person feels satiated after eating a large meal.
b. The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.

Set Point

It is possible to adjust a system’s set point. When this happens, the feedback loop works to maintain the new setting. An example of this is blood pressure: over time, the normal or set point for blood pressure can increase as a result of continued increases in blood pressure. The body no longer recognizes the elevation as abnormal and no attempt is made to return to the lower set point. The result is the maintenance of an elevated blood pressure that can have harmful effects on the body. Medication can lower blood

pressure and lower the set point in the system to a more healthy level. This is called a process of **alteration** of the set point in a feedback loop.

Changes can be made in a group of body organ systems in order to maintain a set point in another system. This is called **acclimatization**. This occurs, for instance, when an animal migrates to a higher altitude than it is accustomed to. In order to adjust to the lower oxygen levels at the new altitude, the body increases the number of red blood cells circulating in the blood to ensure adequate oxygen delivery to the tissues. Another example of acclimatization is animals that have seasonal changes in their coats: a heavier coat in the winter ensures adequate heat retention, and a light coat in summer assists in keeping body temperature from rising to harmful levels.

Note:
Link to Learning

Feedback mechanisms can be understood in terms of driving a race car along a track: watch a short video lesson on positive and negative feedback loops.
https://www.openstaxcollege.org/l/feedback_loops

Homeostasis: Thermoregulation

Body temperature affects body activities. Generally, as body temperature rises, enzyme activity rises as well. For every ten degree centigrade rise in temperature, enzyme activity doubles, up to a point. Body proteins, including enzymes, begin to denature and lose their function with high heat

(around 50^{o}C for mammals). Enzyme activity will decrease by half for every ten degree centigrade drop in temperature, to the point of freezing, with a few exceptions. Some fish can withstand freezing solid and return to normal with thawing.

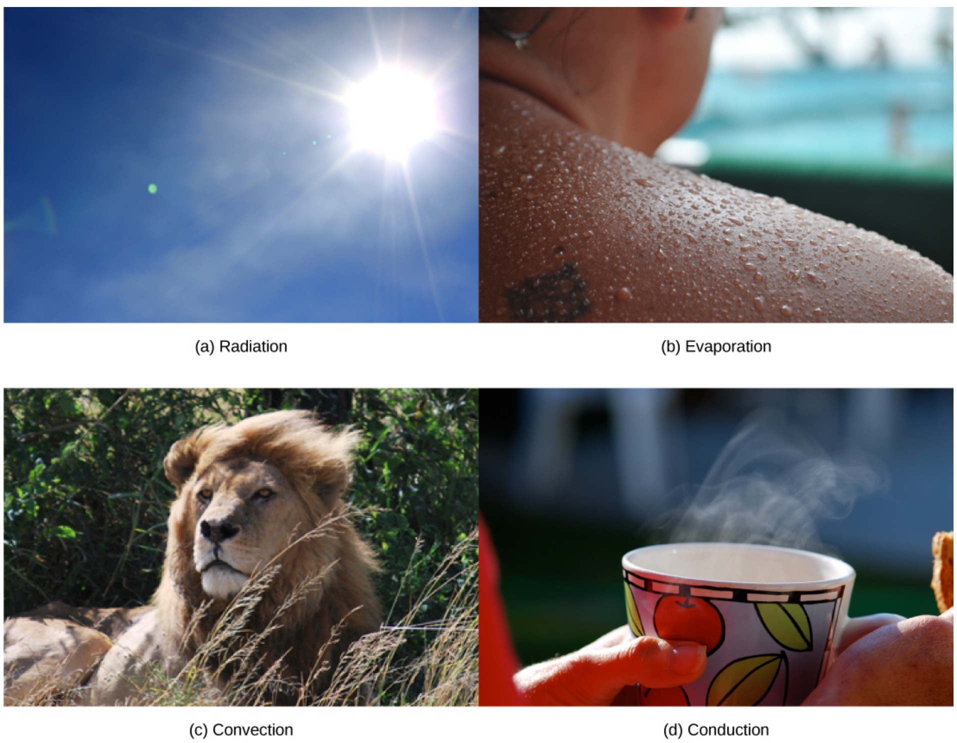
Note:
Link to Learning

Watch this Discovery Channel video on thermoregulation to see illustrations of this process in a variety of animals.
https://www.openstaxcollege.org/l/thermoregulate

Endotherms and Ectotherms

Animals can be divided into two groups: some maintain a constant body temperature in the face of differing environmental temperatures, while others have a body temperature that is the same as their environment and thus varies with the environment. Animals that do not control their body temperature are ectotherms. This group has been called cold-blooded, but the term may not apply to an animal in the desert with a very warm body temperature. In contrast to ectotherms, which rely on external temperatures to set their body temperatures, poikilotherms are animals with constantly varying internal temperatures. An animal that maintains a constant body temperature in the face of environmental changes is called a homeotherm. Endotherms are animals that rely on internal sources for body temperature but which can exhibit extremes in temperature. These animals are able to maintain a level of activity at cooler temperature, which an ectotherm cannot due to differing enzyme levels of activity.

Heat can be exchanged between an animal and its environment through four mechanisms: radiation, evaporation, convection, and conduction ([link]). Radiation is the emission of electromagnetic “heat” waves. Heat comes from the sun in this manner and radiates from dry skin the same way. Heat can be removed with liquid from a surface during evaporation. This occurs when a mammal sweats. Convection currents of air remove heat from the surface of dry skin as the air passes over it. Heat will be conducted from one surface to another during direct contact with the surfaces, such as an animal resting on a warm rock.



Heat can be exchanged by four mechanisms: (a) radiation, (b) evaporation, (c) convection, or (d) conduction. (credit b: modification of work by “Kullez”/Flickr; credit c: modification of work by Chad Rosenthal; credit d: modification of work by “stacey.d”/Flickr)

Heat Conservation and Dissipation

Animals conserve or dissipate heat in a variety of ways. In certain climates, endothermic animals have some form of insulation, such as fur, fat, feathers, or some combination thereof. Animals with thick fur or feathers create an insulating layer of air between their skin and internal organs. Polar bears and seals live and swim in a subfreezing environment and yet maintain a constant, warm, body temperature. The arctic fox, for example, uses its fluffy tail as extra insulation when it curls up to sleep in cold weather. Mammals have a residual effect from shivering and increased muscle activity: arrector pili muscles cause “goose bumps,” causing small hairs to stand up when the individual is cold; this has the intended effect of increasing body temperature. Mammals use layers of fat to achieve the same end. Loss of significant amounts of body fat will compromise an individual’s ability to conserve heat.

Endotherms use their circulatory systems to help maintain body temperature. Vasodilation brings more blood and heat to the body surface, facilitating radiation and evaporative heat loss, which helps to cool the body. Vasoconstriction reduces blood flow in peripheral blood vessels, forcing blood toward the core and the vital organs found there, and conserving heat. Some animals have adaptions to their circulatory system that enable them to transfer heat from arteries to veins, warming blood returning to the heart. This is called a countercurrent heat exchange; it prevents the cold venous blood from cooling the heart and other internal organs. This adaption can be shut down in some animals to prevent overheating the internal organs. The countercurrent adaption is found in many animals, including dolphins, sharks, bony fish, bees, and hummingbirds. In contrast, similar adaptations can help cool endotherms when needed, such as dolphin flukes and elephant ears.

Some ectothermic animals use changes in their behavior to help regulate body temperature. For example, a desert ectothermic animal may simply seek cooler areas during the hottest part of the day in the desert to keep from getting too warm. The same animals may climb onto rocks to capture heat during a cold desert night. Some animals seek water to aid evaporation

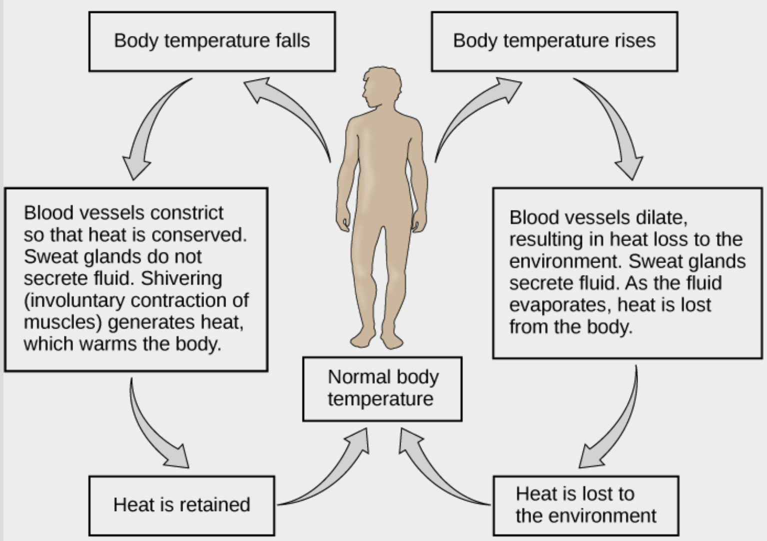
in cooling them, as seen with reptiles. Other ectotherms use group activity such as the activity of bees to warm a hive to survive winter.

Many animals, especially mammals, use metabolic waste heat as a heat source. When muscles are contracted, most of the energy from the ATP used in muscle actions is wasted energy that translates into heat. Severe cold elicits a shivering reflex that generates heat for the body. Many species also have a type of adipose tissue called brown fat that specializes in generating heat.

Neural Control of Thermoregulation

The nervous system is important to **thermoregulation**, as illustrated in [link]. The processes of homeostasis and temperature control are centered in the hypothalamus of the advanced animal brain.

Note:
Art Connection



The body is able to regulate temperature in response to signals from the nervous system.

When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body’s thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?

The hypothalamus maintains the set point for body temperature through reflexes that cause vasodilation and sweating when the body is too warm, or vasoconstriction and shivering when the body is too cold. It responds to chemicals from the body. When a bacterium is destroyed by phagocytic leukocytes, chemicals called endogenous pyrogens are released into the blood. These pyrogens circulate to the hypothalamus and reset the thermostat. This allows the body’s temperature to increase in what is commonly called a fever. An increase in body temperature causes iron to be conserved, which reduces a nutrient needed by bacteria. An increase in body heat also increases the activity of the animal’s enzymes and protective cells while inhibiting the enzymes and activity of the invading microorganisms. Finally, heat itself may also kill the pathogen. A fever that was once thought to be a complication of an infection is now understood to be a normal defense mechanism.

Section Summary

Homeostasis is a dynamic equilibrium that is maintained in body tissues and organs. It is dynamic because it is constantly adjusting to the changes that the systems encounter. It is in equilibrium because body functions are kept within a normal range, with some fluctuations around a set point for the processes.

Art Connections

Exercise:

Problem:

[link] State whether each of the following processes are regulated by a positive feedback loop or a negative feedback loop.

a. A person feels satiated after eating a large meal.
b. The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.

Solution:

[link] Both processes are the result of negative feedback loops. Negative feedback loops, which tend to keep a system at equilibrium, are more common than positive feedback loops.

Exercise:

Problem:

[link] When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body’s thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?

Solution:

[link] Pyrogens increase body temperature by causing the blood vessels to constrict, inducing shivering, and stopping sweat glands from secreting fluid.

Review Questions

Exercise:

Problem:

When faced with a sudden drop in environmental temperature, an endothermic animal will:

a. experience a drop in its body temperature
b. wait to see if it goes lower
c. increase muscle activity to generate heat

d. add fur or fat to increase insulation

Solution:

C

Exercise:

Problem: Which is an example of negative feedback?

a. lowering of blood glucose after a meal
b. blood clotting after an injury
c. lactation during nursing
d. uterine contractions during labor

Solution:

A

Exercise:

Problem:

Which method of heat exchange occurs during direct contact between the source and animal?

a. radiation
b. evaporation
c. convection
d. conduction

Solution:

D

Exercise:

Problem: The body's thermostat is located in the ________.

a. homeostatic receptor
b. hypothalamus
c. medulla
d. vasodilation center

Solution:

B

Free Response

Exercise:

Problem:

Why are negative feedback loops used to control body homeostasis?

Solution:

An adjustment to a change in the internal or external environment requires a change in the direction of the stimulus. A negative feedback loop accomplishes this, while a positive feedback loop would continue the stimulus and result in harm to the animal.

Exercise:

Problem: Why is a fever a "good thing" during a bacterial infection?

Solution:

Mammalian enzymes increase activity to the point of denaturation, increasing the chemical activity of the cells involved. Bacterial enzymes have a specific temperature for their most efficient activity and are inhibited at either higher or lower temperatures. Fever results

in an increase in the destruction of the invading bacteria by increasing the effectiveness of body defenses and an inhibiting bacterial metabolism.

Exercise:

Problem:

How is a condition such as diabetes a good example of the failure of a set point in humans?

Solution:

Diabetes is often associated with a lack in production of insulin. Without insulin, blood glucose levels go up after a meal, but never go back down to normal levels.

Glossary

acclimatization
: alteration in a body system in response to environmental change

alteration
: change of the set point in a homeostatic system

homeostasis
: dynamic equilibrium maintaining appropriate body functions

negative feedback loop
: feedback to a control mechanism that increases or decreases a stimulus instead of maintaining it

positive feedback loop
: feedback to a control mechanism that continues the direction of a stimulus

set point
: midpoint or target point in homeostasis

thermoregulation

regulation of body temperature

Digestive Systems
By the end of this section, you will be able to:

- Explain the processes of digestion and absorption
- Compare and contrast different types of digestive systems
- Explain the specialized functions of the organs involved in processing food in the body
- Describe the ways in which organs work together to digest food and absorb nutrients

Animals obtain their nutrition from the consumption of other organisms. Depending on their diet, animals can be classified into the following categories: plant eaters (herbivores), meat eaters (carnivores), and those that eat both plants and animals (omnivores). The nutrients and macromolecules present in food are not immediately accessible to the cells. There are a number of processes that modify food within the animal body in order to make the nutrients and organic molecules accessible for cellular function. As animals evolved in complexity of form and function, their digestive systems have also evolved to accommodate their various dietary needs.

Herbivores, Omnivores, and Carnivores

Herbivores are animals whose primary food source is plant-based. Examples of herbivores, as shown in [link] include vertebrates like deer, koalas, and some bird species, as well as invertebrates such as crickets and caterpillars. These animals have evolved digestive systems capable of handling large amounts of plant material. Herbivores can be further classified into frugivores (fruit-eaters), granivores (seed eaters), nectivores (nectar feeders), and folivores (leaf eaters).

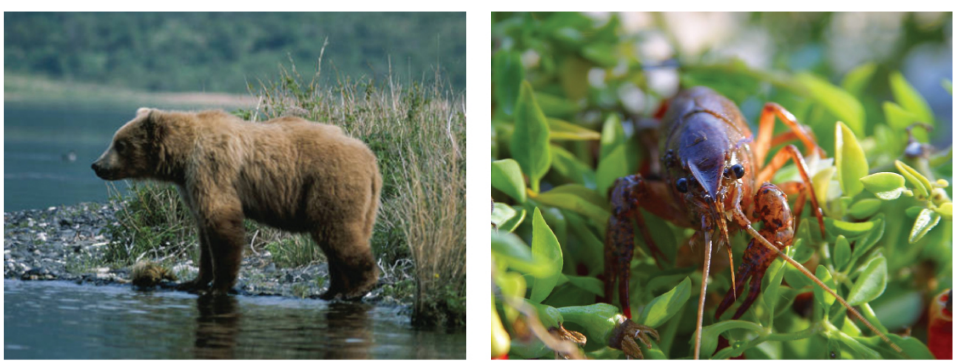
Herbivores, like this (a) mule deer and (b) monarch caterpillar, eat primarily plant material. (credit a: modification of work by Bill Ebbesen; credit b: modification of work by Doug Bowman)

Carnivores are animals that eat other animals. The word carnivore is derived from Latin and literally means “meat eater.” Wild cats such as lions, shown in [link]**a** and tigers are examples of vertebrate carnivores, as are snakes and sharks, while invertebrate carnivores include sea stars, spiders, and ladybugs, shown in [link]**b**. Obligate carnivores are those that rely entirely on animal flesh to obtain their nutrients; examples of obligate carnivores are members of the cat family, such as lions and cheetahs. Facultative carnivores are those that also eat non-animal food in addition to animal food. Note that there is no clear line that differentiates facultative carnivores from omnivores; dogs would be considered facultative carnivores.

(a) (b)

Carnivores like the (a) lion eat primarily meat. The (b) ladybug is also a carnivore that consumes small insects called aphids. (credit a: modification of work by Kevin Pluck; credit b: modification of work by Jon Sullivan)

Omnivores are animals that eat both plant- and animal-derived food. In Latin, omnivore means to eat everything. Humans, bears (shown in [link]**a**), and chickens are example of vertebrate omnivores; invertebrate omnivores include cockroaches and crayfish (shown in [link]**b**).



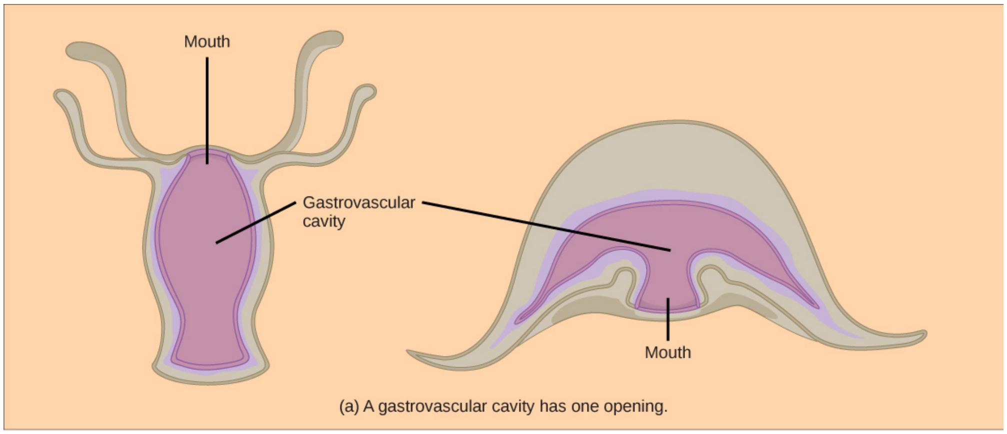
(a) (b)

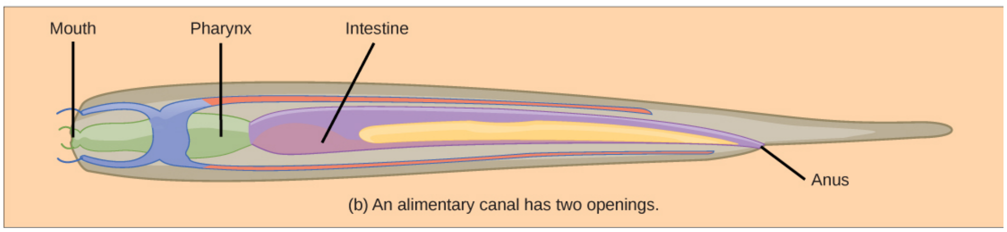
Omnivores like the (a) bear and (b) crayfish eat both plant and animal based food. (credit a: modification of work by Dave Menke; credit b: modification of work by Jon Sullivan)

Invertebrate Digestive Systems

Animals have evolved different types of digestive systems to aid in the digestion of the different foods they consume. The simplest example is that of a **gastrovascular cavity** and is found in organisms with only one opening for digestion. Platyhelminthes (flatworms), Ctenophora (comb jellies), and Cnidaria (coral, jelly fish, and sea anemones) use this type of digestion. Gastrovascular cavities, as shown in [link]**a**, are typically a blind tube or cavity with only one opening, the “mouth”, which also serves as an “anus”. Ingested material enters the mouth and passes through a hollow, tubular cavity. Cells within the cavity secrete digestive enzymes that break down the food. The food particles are engulfed by the cells lining the gastrovascular cavity.

The **alimentary canal**, shown in [link]**b**, is a more advanced system: it consists of one tube with a mouth at one end and an anus at the other. Earthworms are an example of an animal with an alimentary canal. Once the food is ingested through the mouth, it passes through the esophagus and is stored in an organ called the crop; then it passes into the gizzard where it is churned and digested. From the gizzard, the food passes through the intestine, the nutrients are absorbed, and the waste is eliminated as feces, called castings, through the anus.





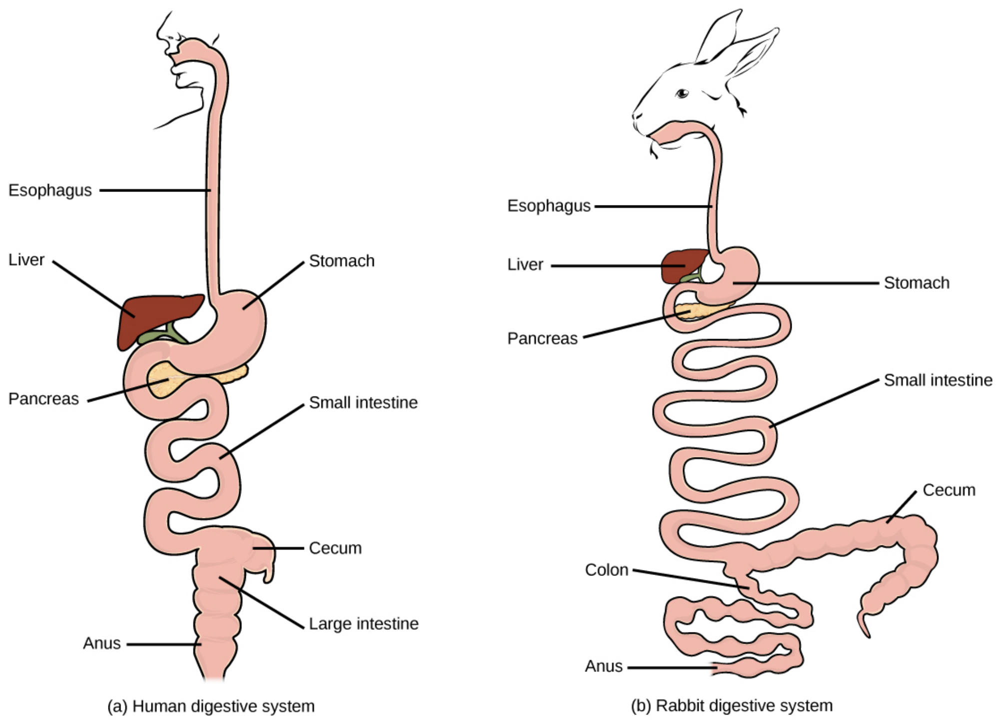
(a) A gastrovascular cavity has a single opening through which food is ingested and waste is excreted, as shown in this hydra and in this jellyfish medusa. (b) An alimentary canal has two openings: a mouth for ingesting food, and an anus for eliminating waste, as shown in this nematode.

Vertebrate Digestive Systems

Vertebrates have evolved more complex digestive systems to adapt to their dietary needs. Some animals have a single stomach, while others have multi-chambered stomachs. Birds have developed a digestive system adapted to eating unmasticated food.

Monogastric: Single-chambered Stomach

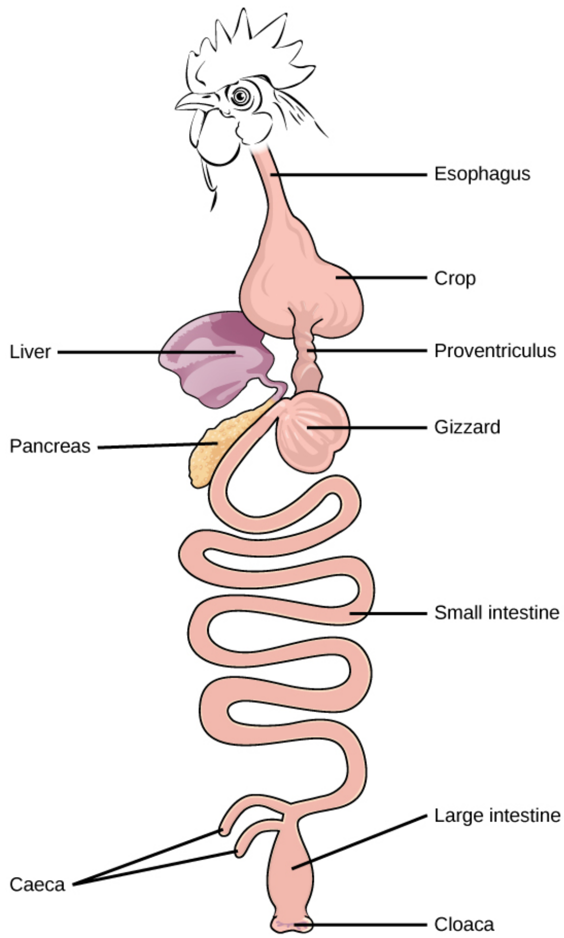
As the word **monogastric** suggests, this type of digestive system consists of one (“mono”) stomach chamber (“gastric”). Humans and many animals have a monogastric digestive system as illustrated in [link]**ab**. The process of digestion begins with the mouth and the intake of food. The teeth play an important role in masticating (chewing) or physically breaking down food into smaller particles. The enzymes present in saliva also begin to chemically break down food. The esophagus is a long tube that connects the mouth to the stomach. Using peristalsis, or wave-like smooth muscle contractions, the muscles of the esophagus push the food towards the stomach. In order to speed up the actions of enzymes in the stomach, the stomach is an extremely acidic environment, with a pH between 1.5 and 2.5. The gastric juices, which include enzymes in the stomach, act on the food particles and continue the process of digestion. Further breakdown of food takes place in the small intestine where enzymes produced by the liver, the small intestine, and the pancreas continue the process of digestion. The nutrients are absorbed into the blood stream across the epithelial cells lining the walls of the small intestines. The waste material travels on to the large intestine where water is absorbed and the drier waste material is compacted into feces; it is stored until it is excreted through the rectum.



(a) Humans and herbivores, such as the (b) rabbit, have a monogastric digestive system. However, in the rabbit the small intestine and cecum are enlarged to allow more time to digest plant material. The enlarged organ provides more surface area for absorption of nutrients. Rabbits digest their food twice: the first time food passes through the digestive system, it collects in the cecum, and then it passes as soft feces called cecotrophes. The rabbit re-ingests these cecotrophes to further digest them.

Avian

Birds face special challenges when it comes to obtaining nutrition from food. They do not have teeth and so their digestive system, shown in [link], must be able to process un-masticated food. Birds have evolved a variety of beak types that reflect the vast variety in their diet, ranging from seeds and insects to fruits and nuts. Because most birds fly, their metabolic rates are high in order to efficiently process food and keep their body weight low. The stomach of birds has two chambers: the **proventriculus**, where gastric juices are produced to digest the food before it enters the stomach, and the **gizzard**, where the food is stored, soaked, and mechanically ground. The undigested material forms food pellets that are sometimes regurgitated. Most of the chemical digestion and absorption happens in the intestine and the waste is excreted through the cloaca.



The avian esophagus has a pouch, called a crop, which stores food. Food passes from the crop to the first of two stomachs, called the proventriculus, which contains digestive juices that break down food. From the proventriculus, the food enters the second stomach, called the gizzard, which grinds food. Some birds swallow stones or grit, which are stored in the gizzard, to aid the grinding process. Birds do not have separate

openings to excrete urine and feces. Instead, uric acid from the kidneys is secreted into the large intestine and combined with waste from the digestive process. This waste is excreted through an opening called the cloaca.

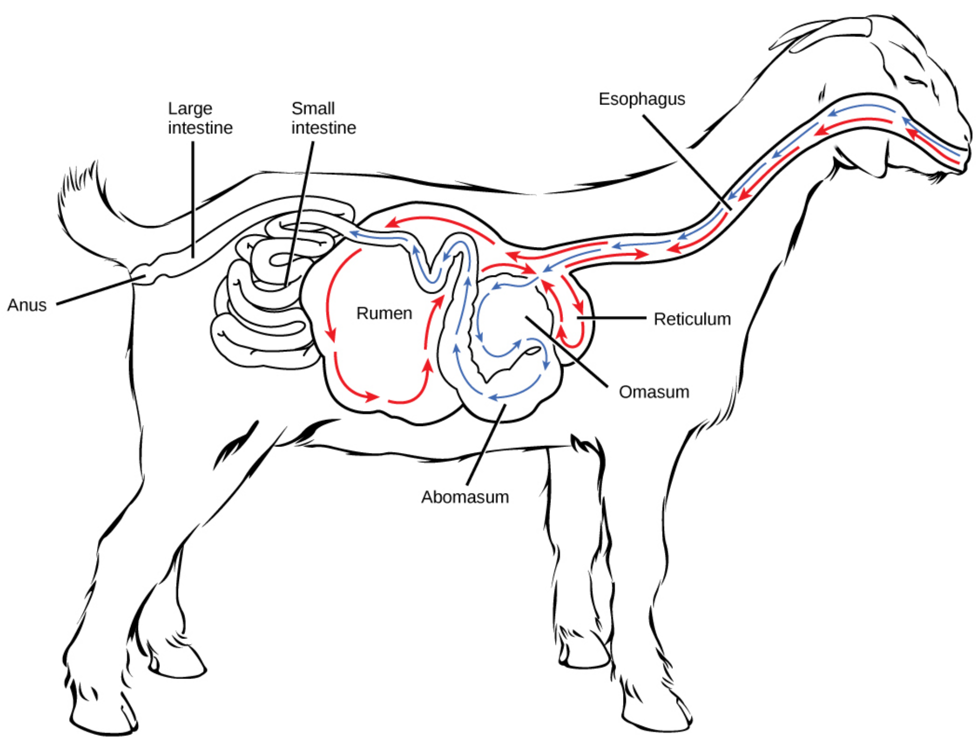
Note:
Evolution Connection
Avian Adaptations
Birds have a highly efficient, simplified digestive system. Recent fossil evidence has shown that the evolutionary divergence of birds from other land animals was characterized by streamlining and simplifying the digestive system. Unlike many other animals, birds do not have teeth to chew their food. In place of lips, they have sharp pointy beaks. The horny beak, lack of jaws, and the smaller tongue of the birds can be traced back to their dinosaur ancestors. The emergence of these changes seems to coincide with the inclusion of seeds in the bird diet. Seed-eating birds have beaks that are shaped for grabbing seeds and the two-compartment stomach allows for delegation of tasks. Since birds need to remain light in order to fly, their metabolic rates are very high, which means they digest their food very quickly and need to eat often. Contrast this with the ruminants, where the digestion of plant matter takes a very long time.

Ruminants

Ruminants are mainly herbivores like cows, sheep, and goats, whose entire diet consists of eating large amounts of **roughage** or fiber. They have evolved digestive systems that help them digest vast amounts of cellulose. An interesting feature of the ruminants’ mouth is that they do not have

upper incisor teeth. They use their lower teeth, tongue and lips to tear and chew their food. From the mouth, the food travels to the esophagus and on to the stomach.

To help digest the large amount of plant material, the stomach of the ruminants is a multi-chambered organ, as illustrated in [link]. The four compartments of the stomach are called the rumen, reticulum, omasum, and abomasum. These chambers contain many microbes that break down cellulose and ferment ingested food. The abomasum is the “true” stomach and is the equivalent of the monogastric stomach chamber where gastric juices are secreted. The four-compartment gastric chamber provides larger space and the microbial support necessary to digest plant material in ruminants. The fermentation process produces large amounts of gas in the stomach chamber, which must be eliminated. As in other animals, the small intestine plays an important role in nutrient absorption, and the large intestine helps in the elimination of waste.



Ruminant animals, such as goats and cows, have four stomachs. The first two stomachs, the rumen and the

reticulum, contain prokaryotes and protists that are able to digest cellulose fiber. The ruminant regurgitates cud from the reticulum, chews it, and swallows it into a third stomach, the omasum, which removes water. The cud then passes onto the fourth stomach, the abomasum, where it is digested by enzymes produced by the ruminant.

Pseudo-ruminants

Some animals, such as camels and alpacas, are pseudo-ruminants. They eat a lot of plant material and roughage. Digesting plant material is not easy because plant cell walls contain the polymeric sugar molecule cellulose. The digestive enzymes of these animals cannot break down cellulose, but microorganisms present in the digestive system can. Therefore, the digestive system must be able to handle large amounts of roughage and break down the cellulose. Pseudo-ruminants have a three-chamber stomach in the digestive system. However, their cecum—a pouched organ at the beginning of the large intestine containing many microorganisms that are necessary for the digestion of plant materials—is large and is the site where the roughage is fermented and digested. These animals do not have a rumen but have an omasum, abomasum, and reticulum.

Parts of the Digestive System

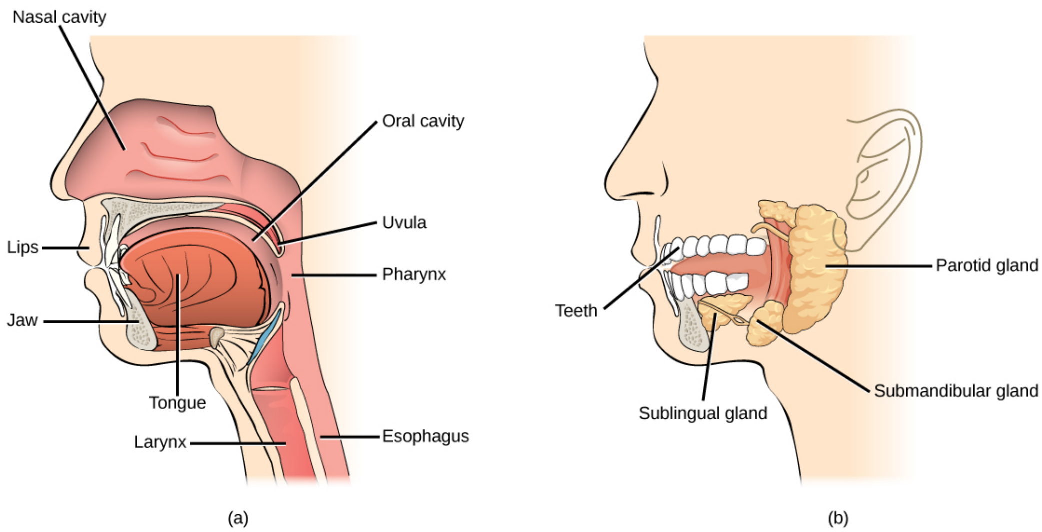
The vertebrate digestive system is designed to facilitate the transformation of food matter into the nutrient components that sustain organisms.

Oral Cavity

The oral cavity, or mouth, is the point of entry of food into the digestive system, illustrated in [link]. The food consumed is broken into smaller

particles by mastication, the chewing action of the teeth. All mammals have teeth and can chew their food.

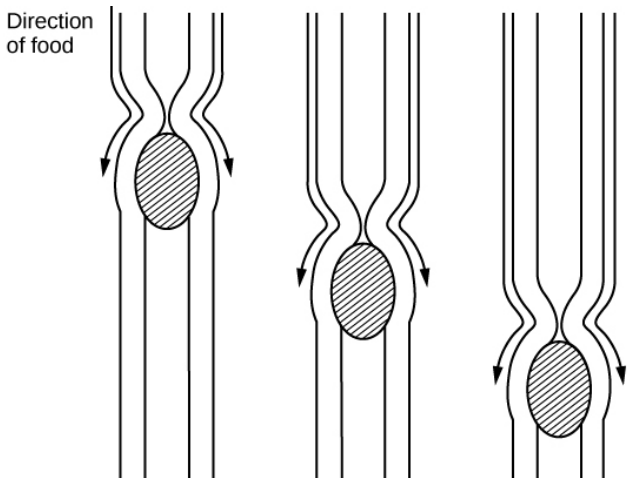
The extensive chemical process of digestion begins in the mouth. As food is being chewed, saliva, produced by the salivary glands, mixes with the food. Saliva is a watery substance produced in the mouths of many animals. There are three major glands that secrete saliva—the parotid, the submandibular, and the sublingual. Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains immunoglobulins and lysozymes, which have antibacterial action to reduce tooth decay by inhibiting growth of some bacteria. Saliva also contains an enzyme called **salivary amylase** that begins the process of converting starches in the food into a disaccharide called maltose. Another enzyme called **lipase** is produced by the cells in the tongue. Lipases are a class of enzymes that can break down triglycerides. The lingual lipase begins the breakdown of fat components in the food. The chewing and wetting action provided by the teeth and saliva prepare the food into a mass called the **bolus** for swallowing. The tongue helps in swallowing—moving the bolus from the mouth into the pharynx. The pharynx opens to two passageways: the trachea, which leads to the lungs, and the esophagus, which leads to the stomach. The trachea has an opening called the glottis, which is covered by a cartilaginous flap called the epiglottis. When swallowing, the epiglottis closes the glottis and food passes into the esophagus and not the trachea. This arrangement allows food to be kept out of the trachea.



Digestion of food begins in the (a) oral cavity. Food is masticated by teeth and moistened by saliva secreted from the (b) salivary glands. Enzymes in the saliva begin to digest starches and fats. With the help of the tongue, the resulting bolus is moved into the esophagus by swallowing. (credit: modification of work by the National Cancer Institute)

Esophagus

The **esophagus** is a tubular organ that connects the mouth to the stomach. The chewed and softened food passes through the esophagus after being swallowed. The smooth muscles of the esophagus undergo a series of wave like movements called **peristalsis** that push the food toward the stomach, as illustrated in [link]. The peristalsis wave is unidirectional—it moves food from the mouth to the stomach, and reverse movement is not possible. The peristaltic movement of the esophagus is an involuntary reflex; it takes place in response to the act of swallowing.



The esophagus transfers food from the mouth to the stomach through peristaltic movements.

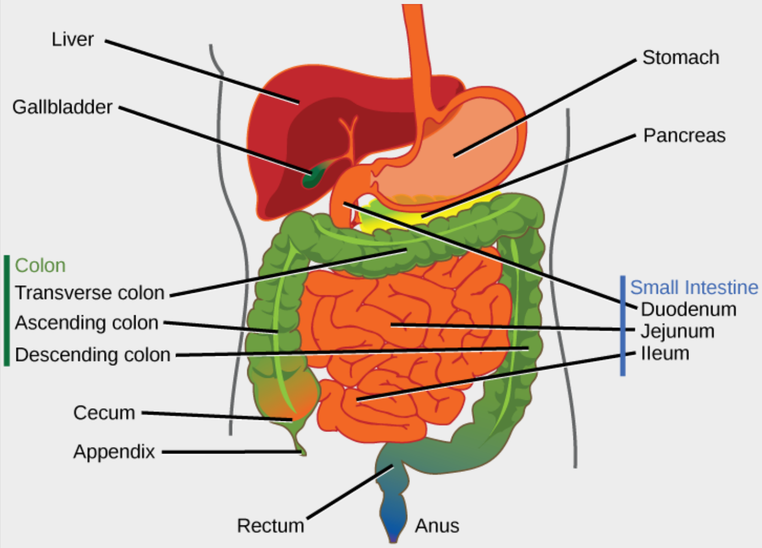
A ring-like muscle called a **sphincter** forms valves in the digestive system. The gastro-esophageal sphincter is located at the stomach end of the esophagus. In response to swallowing and the pressure exerted by the bolus of food, this sphincter opens, and the bolus enters the stomach. When there is no swallowing action, this sphincter is shut and prevents the contents of the stomach from traveling up the esophagus. Many animals have a true sphincter; however, in humans, there is no true sphincter, but the esophagus remains closed when there is no swallowing action. Acid reflux or "heartburn" occurs when the acidic digestive juices escape into the esophagus.

Stomach

A large part of digestion occurs in the stomach, shown in [link]. The **stomach** is a saclike organ that secretes gastric digestive juices. The pH in the stomach is between 1.5 and 2.5. This highly acidic environment is required for the chemical breakdown of food and the extraction of nutrients. When empty, the stomach is a rather small organ; however, it can expand to

up to 20 times its resting size when filled with food. This characteristic is particularly useful for animals that need to eat when food is available.

Note:
Art Connection



The human stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

Which of the following statements about the digestive system is false?

a. Chyme is a mixture of food and digestive juices that is produced in the stomach.
b. Food enters the large intestine before the small intestine.
c. In the small intestine, chyme mixes with bile, which emulsifies fats.
d. The stomach is separated from the small intestine by the pyloric sphincter.

The stomach is also the major site for protein digestion in animals other than ruminants. Protein digestion is mediated by an enzyme called pepsin in the stomach chamber. **Pepsin** is secreted by the chief cells in the stomach in an inactive form called **pepsinogen**. Pepsin breaks peptide bonds and cleaves proteins into smaller polypeptides; it also helps activate more pepsinogen, starting a positive feedback mechanism that generates more pepsin. Another cell type—parietal cells—secrete hydrogen and chloride ions, which combine in the lumen to form hydrochloric acid, the primary acidic component of the stomach juices. Hydrochloric acid helps to convert the inactive pepsinogen to pepsin. The highly acidic environment also kills many microorganisms in the food and, combined with the action of the enzyme pepsin, results in the hydrolysis of protein in the food. Chemical digestion is facilitated by the churning action of the stomach. Contraction and relaxation of smooth muscles mixes the stomach contents about every 20 minutes. The partially digested food and gastric juice mixture is called **chyme**. Chyme passes from the stomach to the small intestine. Further protein digestion takes place in the small intestine. Gastric emptying occurs within two to six hours after a meal. Only a small amount of chyme is released into the small intestine at a time. The movement of chyme from the stomach into the small intestine is regulated by the pyloric sphincter.

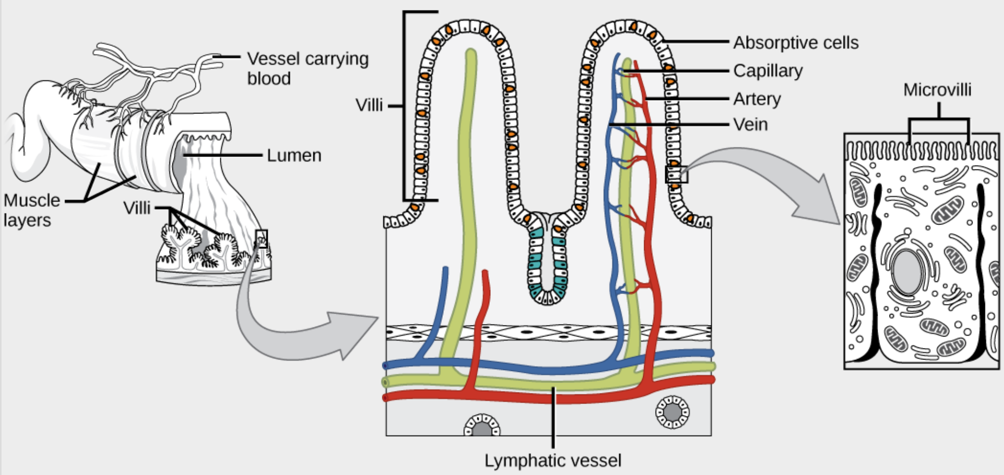
When digesting protein and some fats, the stomach lining must be protected from getting digested by pepsin. There are two points to consider when describing how the stomach lining is protected. First, as previously mentioned, the enzyme pepsin is synthesized in the inactive form. This protects the chief cells, because pepsinogen does not have the same enzyme functionality of pepsin. Second, the stomach has a thick mucus lining that protects the underlying tissue from the action of the digestive juices. When this mucus lining is ruptured, ulcers can form in the stomach. Ulcers are open wounds in or on an organ caused by bacteria (*Helicobacter pylori*) when the mucus lining is ruptured and fails to reform.

Small Intestine

Chyme moves from the stomach to the small intestine. The **small intestine** is the organ where the digestion of protein, fats, and carbohydrates is

completed. The small intestine is a long tube-like organ with a highly folded surface containing finger-like projections called the **villi**. The apical surface of each villus has many microscopic projections called microvilli. These structures, illustrated in [link], are lined with epithelial cells on the luminal side and allow for the nutrients to be absorbed from the digested food and absorbed into the blood stream on the other side. The villi and microvilli, with their many folds, increase the surface area of the intestine and increase absorption efficiency of the nutrients. Absorbed nutrients in the blood are carried into the hepatic portal vein, which leads to the liver. There, the liver regulates the distribution of nutrients to the rest of the body and removes toxic substances, including drugs, alcohol, and some pathogens.

Note:
Art Connection



Villi are folds on the small intestine lining that increase the surface area to facilitate the absorption of nutrients.

Which of the following statements about the small intestine is false?

a. Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.

b. The inside of the small intestine has many folds, called villi.
c. Microvilli are lined with blood vessels as well as lymphatic vessels.
d. The inside of the small intestine is called the lumen.

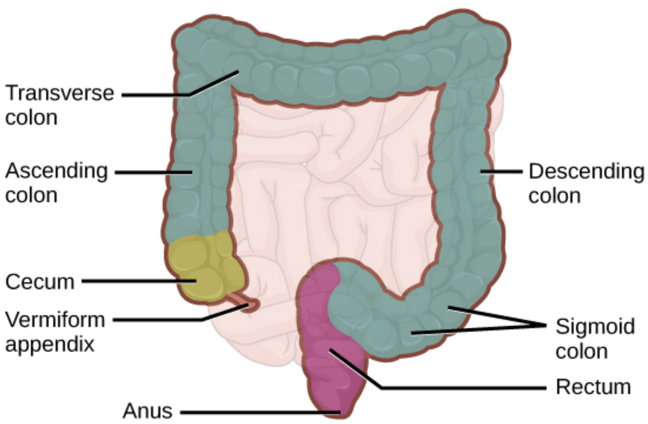
The human small intestine is over 6m long and is divided into three parts: the duodenum, the jejunum, and the ileum. The “C-shaped,” fixed part of the small intestine is called the **duodenum** and is shown in [link]. The duodenum is separated from the stomach by the pyloric sphincter which opens to allow chyme to move from the stomach to the duodenum. In the duodenum, chyme is mixed with pancreatic juices in an alkaline solution rich in bicarbonate that neutralizes the acidity of chyme and acts as a buffer. Pancreatic juices also contain several digestive enzymes. Digestive juices from the pancreas, liver, and gallbladder, as well as from gland cells of the intestinal wall itself, enter the duodenum. **Bile** is produced in the liver and stored and concentrated in the gallbladder. Bile contains bile salts which emulsify lipids while the pancreas produces enzymes that catabolize starches, disaccharides, proteins, and fats. These digestive juices break down the food particles in the chyme into glucose, triglycerides, and amino acids. Some chemical digestion of food takes place in the duodenum. Absorption of fatty acids also takes place in the duodenum.

The second part of the small intestine is called the **jejunum**, shown in [link]. Here, hydrolysis of nutrients is continued while most of the carbohydrates and amino acids are absorbed through the intestinal lining. The bulk of chemical digestion and nutrient absorption occurs in the jejunum.

The **ileum**, also illustrated in [link] is the last part of the small intestine and here the bile salts and vitamins are absorbed into blood stream. The undigested food is sent to the colon from the ileum via peristaltic movements of the muscle. The ileum ends and the large intestine begins at the ileocecal valve. The vermiform, “worm-like,” appendix is located at the ileocecal valve. The appendix of humans secretes no enzymes and has an insignificant role in immunity.

Large Intestine

The **large intestine**, illustrated in [link], reabsorbs the water from the undigested food material and processes the waste material. The human large intestine is much smaller in length compared to the small intestine but larger in diameter. It has three parts: the cecum, the colon, and the rectum. The cecum joins the ileum to the colon and is the receiving pouch for the waste matter. The colon is home to many bacteria or “intestinal flora” that aid in the digestive processes. The colon can be divided into four regions, the ascending colon, the transverse colon, the descending colon and the sigmoid colon. The main functions of the colon are to extract the water and mineral salts from undigested food, and to store waste material. Carnivorous mammals have a shorter large intestine compared to herbivorous mammals due to their diet.



The large intestine reabsorbs water from undigested food and stores waste material until it is eliminated.

Rectum and Anus

The **rectum** is the terminal end of the large intestine, as shown in [link]. The primary role of the rectum is to store the feces until defecation. The feces are propelled using peristaltic movements during elimination. The **anus** is an opening at the far-end of the digestive tract and is the exit point for the waste material. Two sphincters between the rectum and anus control elimination: the inner sphincter is involuntary and the outer sphincter is voluntary.

Accessory Organs

The organs discussed above are the organs of the digestive tract through which food passes. Accessory organs are organs that add secretions (enzymes) that catabolize food into nutrients. Accessory organs include salivary glands, the liver, the pancreas, and the gallbladder. The liver, pancreas, and gallbladder are regulated by hormones in response to the food consumed.

The **liver** is the largest internal organ in humans and it plays a very important role in digestion of fats and detoxifying blood. The liver produces bile, a digestive juice that is required for the breakdown of fatty components of the food in the duodenum. The liver also processes the vitamins and fats and synthesizes many plasma proteins.

The **pancreas** is another important gland that secretes digestive juices. The chyme produced from the stomach is highly acidic in nature; the pancreatic juices contain high levels of bicarbonate, an alkali that neutralizes the acidic chyme. Additionally, the pancreatic juices contain a large variety of enzymes that are required for the digestion of protein and carbohydrates.

The **gallbladder** is a small organ that aids the liver by storing bile and concentrating bile salts. When chyme containing fatty acids enters the duodenum, the bile is secreted from the gallbladder into the duodenum.

Section Summary

Different animals have evolved different types of digestive systems specialized to meet their dietary needs. Humans and many other animals have monogastric digestive systems with a single-chambered stomach. Birds have evolved a digestive system that includes a gizzard where the food is crushed into smaller pieces. This compensates for their inability to masticate. Ruminants that consume large amounts of plant material have a multi-chambered stomach that digests roughage. Pseudo-ruminants have similar digestive processes as ruminants but do not have the four-compartment stomach. Processing food involves ingestion (eating), digestion (mechanical and enzymatic breakdown of large molecules), absorption (cellular uptake of nutrients), and elimination (removal of undigested waste as feces).

Many organs work together to digest food and absorb nutrients. The mouth is the point of ingestion and the location where both mechanical and chemical breakdown of food begins. Saliva contains an enzyme called amylase that breaks down carbohydrates. The food bolus travels through the esophagus by peristaltic movements to the stomach. The stomach has an extremely acidic environment. An enzyme called pepsin digests protein in the stomach. Further digestion and absorption take place in the small intestine. The large intestine reabsorbs water from the undigested food and stores waste until elimination.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the digestive system is false?

a. Chyme is a mixture of food and digestive juices that is produced in the stomach.
b. Food enters the large intestine before the small intestine.
c. In the small intestine, chyme mixes with bile, which emulsifies fats.

d. The stomach is separated from the small intestine by the pyloric sphincter.

Solution:

[link] B

Exercise:

Problem:

[link] Which of the following statements about the small intestine is false?

a. Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.
b. The inside of the small intestine has many folds, called villi.
c. Microvilli are lined with blood vessels as well as lymphatic vessels.
d. The inside of the small intestine is called the lumen.

Solution:

[link] C

Review Questions

Exercise:

Problem: Which of the following is a pseudo-ruminant?

a. cow
b. pig
c. crow
d. horse

Solution:

D

Exercise:

Problem: Which of the following statements is untrue?

a. Roughage takes a long time to digest.
b. Birds eat large quantities at one time so that they can fly long distances.
c. Cows do not have upper teeth.
d. In pseudo-ruminants, roughage is digested in the cecum.

Solution:

B

Exercise:

Problem: The acidic nature of chyme is neutralized by _________.

a. potassium hydroxide
b. sodium hydroxide
c. bicarbonates
d. vinegar

Solution:

C

Exercise:

Problem:

The digestive juices from the liver are delivered to the _________.

a. stomach
b. liver
c. duodenum
d. colon

Solution:

C

Free Response

Exercise:

Problem:

How does the polygastric digestive system aid in digesting roughage?

Solution:

Animals with a polygastric digestive system have a multi-chambered stomach. The four compartments of the stomach are called the rumen, reticulum, omasum, and abomasum. These chambers contain many microbes that break down the cellulose and ferment the ingested food. The abomasum is the "true" stomach and is the equivalent of a monogastric stomach chamber where gastric juices are secreted. The four-compartment gastric chamber provides larger space and the microbial support necessary for ruminants to digest plant material.

Exercise:

Problem: How do birds digest their food in the absence of teeth?

Solution:

Birds have a stomach chamber called a gizzard. Here, the food is stored, soaked, and ground into finer particles, often using pebbles.

Once this process is complete, the digestive juices take over in the proventriculus and continue the digestive process.

Exercise:

Problem: What is the role of the accessory organs in digestion?

Solution:

Accessory organs play an important role in producing and delivering digestive juices to the intestine during digestion and absorption. Specifically, the salivary glands, liver, pancreas, and gallbladder play important roles. Malfunction of any of these organs can lead to disease states.

Exercise:

Problem: Explain how the villi and microvilli aid in absorption.

Solution:

The villi and microvilli are folds on the surface of the small intestine. These folds increase the surface area of the intestine and provide more area for the absorption of nutrients.

Glossary

alimentary canal
: tubular digestive system with a mouth and anus

anus
: exit point for waste material

bile
: digestive juice produced by the liver; important for digestion of lipids

bolus

mass of food resulting from chewing action and wetting by saliva

carnivore
animal that consumes animal flesh

chyme
mixture of partially digested food and stomach juices

duodenum
first part of the small intestine where a large part of digestion of carbohydrates and fats occurs

esophagus
tubular organ that connects the mouth to the stomach

gallbladder
organ that stores and concentrates bile

gastrovascular cavity
digestive system consisting of a single opening

gizzard
muscular organ that grinds food

herbivore
animal that consumes strictly plant diet

ileum
last part of the small intestine; connects the small intestine to the large intestine; important for absorption of B-12

jejunum
second part of the small intestine

large intestine
digestive system organ that reabsorbs water from undigested material and processes waste matter

lipase

enzyme that chemically breaks down lipids

liver
organ that produces bile for digestion and processes vitamins and lipids

monogastric
digestive system that consists of a single-chambered stomach

omnivore
animal that consumes both plants and animals

pancreas
gland that secretes digestive juices

pepsin
enzyme found in the stomach whose main role is protein digestion

pepsinogen
inactive form of pepsin

peristalsis
wave-like movements of muscle tissue

proventriculus
glandular part of a bird’s stomach

rectum
area of the body where feces is stored until elimination

roughage
component of food that is low in energy and high in fiber

ruminant
animal with a stomach divided into four compartments

salivary amylase
enzyme found in saliva, which converts carbohydrates to maltose

small intestine

organ where digestion of protein, fats, and carbohydrates is completed

sphincter

band of muscle that controls movement of materials throughout the digestive tract

stomach

saclike organ containing acidic digestive juices

villi

folds on the inner surface of the small intestine whose role is to increase absorption area

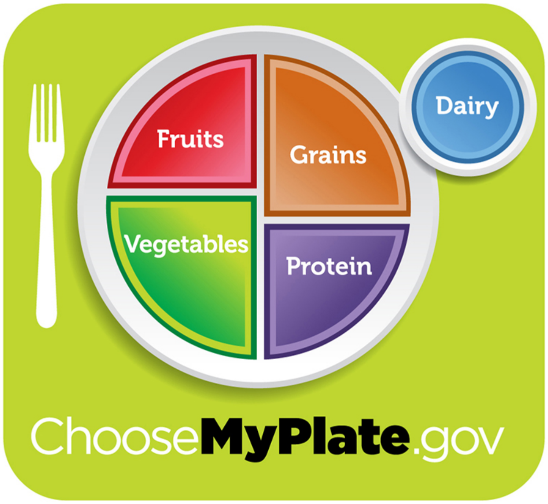
Nutrition and Energy Production
By the end of this section, you will be able to:

- Explain why an animal's diet should be balanced and meet the needs of the body
- Define the primary components of food
- Describe the essential nutrients required for cellular function that cannot be synthesized by the animal body
- Explain how energy is produced through diet and digestion
- Describe how excess carbohydrates and energy are stored in the body

Given the diversity of animal life on our planet, it is not surprising that the animal diet would also vary substantially. The animal diet is the source of materials needed for building DNA and other complex molecules needed for growth, maintenance, and reproduction; collectively these processes are called biosynthesis. The diet is also the source of materials for ATP production in the cells. The diet must be balanced to provide the minerals and vitamins that are required for cellular function.

Food Requirements

What are the fundamental requirements of the animal diet? The animal diet should be well balanced and provide nutrients required for bodily function and the minerals and vitamins required for maintaining structure and regulation necessary for good health and reproductive capability. These requirements for a human are illustrated graphically in [link]



For humans, a balanced diet includes fruits, vegetables, grains, and protein. (credit: USDA)

Note:
Link to Learning

The first step in ensuring that you are meeting the food requirements of your body is an awareness of the food groups and the nutrients they provide. To learn more about each food group and the recommended daily amounts, explore this interactive site by the United States Department of Agriculture.

Note:
Everyday Connection
Let’s Move! Campaign
Obesity is a growing epidemic and the rate of obesity among children is rapidly rising in the United States. To combat childhood obesity and ensure that children get a healthy start in life, first lady Michelle Obama has launched the Let’s Move! campaign. The goal of this campaign is to educate parents and caregivers on providing healthy nutrition and encouraging active lifestyles to future generations. This program aims to involve the entire community, including parents, teachers, and healthcare providers to ensure that children have access to healthy foods—more fruits, vegetables, and whole grains—and consume fewer calories from processed foods. Another goal is to ensure that children get physical activity. With the increase in television viewing and stationary pursuits such as video games, sedentary lifestyles have become the norm. Learn more at www.letsmove.gov.

Organic Precursors

The organic molecules required for building cellular material and tissues must come from food. Carbohydrates or sugars are the primary source of organic carbons in the animal body. During digestion, digestible carbohydrates are ultimately broken down into glucose and used to provide energy through metabolic pathways. Complex carbohydrates, including polysaccharides, can be broken down into glucose through biochemical modification; however, humans do not produce the enzyme cellulase and lack the ability to derive glucose from the polysaccharide cellulose. In humans, these molecules provide the fiber required for moving waste through the large intestine and a healthy colon. The intestinal flora in the human gut are able to extract some nutrition from these plant fibers. The excess sugars in the body are converted into glycogen and stored in the liver and muscles for later use. Glycogen stores are used to fuel prolonged exertions, such as long-distance running, and to provide energy during food shortage. Excess glycogen can be converted to fats, which are stored in the lower layer of the skin of mammals for insulation and energy storage. Excess digestible carbohydrates are stored by mammals in order to survive famine and aid in mobility.

Another important requirement is that of nitrogen. Protein catabolism provides a source of organic nitrogen. Amino acids are the building blocks of proteins and protein breakdown provides amino acids that are used for cellular function. The carbon and nitrogen derived from these become the building block for nucleotides, nucleic acids, proteins, cells, and tissues. Excess nitrogen must be excreted as it is toxic. Fats add flavor to food and promote a sense of satiety or fullness. Fatty foods are also significant sources of energy because one gram of fat contains nine calories. Fats are required in the diet to aid the absorption of fat-soluble vitamins and the production of fat-soluble hormones.

Essential Nutrients

While the animal body can synthesize many of the molecules required for function from the organic precursors, there are some nutrients that need to be consumed from food. These nutrients are termed **essential nutrients**, meaning they must be eaten, and the body cannot produce them.

The omega-3 alpha-linolenic acid and the omega-6 linoleic acid are essential fatty acids needed to make some membrane phospholipids. **Vitamins** are another class of essential organic molecules that are required in small quantities for many enzymes to function and, for this reason, are considered to be co-enzymes. Absence or low levels of vitamins can have a dramatic effect on health, as outlined in [link] and [link]. Both fat-soluble and water-soluble vitamins must be obtained from food. **Minerals**, listed in [link], are inorganic essential nutrients that must be obtained from food. Among their many functions, minerals help in structure and regulation and are considered co-factors. Certain amino acids also must be procured from food and cannot be synthesized by the body. These amino acids are the “essential” amino acids. The human body can synthesize only 11 of the 20 required amino acids; the rest must be obtained from food. The essential amino acids are listed in [link].

Water-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**
Vitamin B_1 (Thiamine)	Needed by the body to process lipids, proteins, and carbohydrates Coenzyme removes CO_2 from organic compounds	Muscle weakness, Beriberi: reduced heart function, CNS problems	Milk, meat, dried beans, whole grains
Vitamin B_2 (Riboflavin)	Takes an active role in metabolism, aiding in the conversion of food to energy (FAD and FMN)	Cracks or sores on the outer surface of the lips (cheliosis); inflammation and redness of the tongue; moist, scaly skin inflammation (seborrheic dermatitis)	Meat, eggs, enriched grains, vegetables

Water-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**
Vitamin B_3 (Niacin)	Used by the body to release energy from carbohydrates and to process alcohol; required for the synthesis of sex hormones; component of coenzyme NAD^+ and $NADP^+$	Pellagra, which can result in dermatitis, diarrhea, dementia, and death	Meat, eggs, grains, nuts, potatoes
Vitamin B_5 (Pantothenic acid)	Assists in producing energy from foods (lipids, in particular); component of coenzyme A	Fatigue, poor coordination, retarded growth, numbness, tingling of hands and feet	Meat, whole grains, milk, fruits, vegetables
Vitamin B_6 (Pyridoxine)	The principal vitamin for processing amino acids and lipids; also helps convert nutrients into energy	Irritability, depression, confusion, mouth sores or ulcers, anemia, muscular twitching	Meat, dairy products, whole grains, orange juice

Water-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**
Vitamin B_7 (Biotin)	Used in energy and amino acid metabolism, fat synthesis, and fat breakdown; helps the body use blood sugar	Hair loss, dermatitis, depression, numbness and tingling in the extremities; neuromuscular disorders	Meat, eggs, legumes and other vegetables
Vitamin B_9 (Folic acid)	Assists the normal development of cells, especially during fetal development; helps metabolize nucleic and amino acids	Deficiency during pregnancy is associated with birth defects, such as neural tube defects and anemia	Leafy green vegetables, whole wheat, fruits, nuts, legumes
Vitamin B_{12} (Cobalamin)	Maintains healthy nervous system and assists with blood cell formation; coenzyme in nucleic acid metabolism	Anemia, neurological disorders, numbness, loss of balance	Meat, eggs, animal products

Water-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**
Vitamin C (Ascorbic acid)	Helps maintain connective tissue: bone, cartilage, and dentin; boosts the immune system	Scurvy, which results in bleeding, hair and tooth loss; joint pain and swelling; delayed wound healing	Citrus fruits, broccoli, tomatoes, red sweet bell peppers

Fat-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**

Fat-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**
Vitamin A (Retinol)	Critical to the development of bones, teeth, and skin; helps maintain eyesight, enhances the immune system, fetal development, gene expression	Night-blindness, skin disorders, impaired immunity	Dark green leafy vegetables, yellow-orange vegetables fruits, milk, butter
Vitamin D	Critical for calcium absorption for bone development and strength; maintains a stable nervous system; maintains a normal and strong heartbeat; helps in blood clotting	Rickets, osteomalacia, immunity	Cod liver oil, milk, egg yolk

Fat-soluble Essential Vitamins			
Vitamin	**Function**	**Deficiencies Can Lead To**	**Sources**
Vitamin E (Tocopherol)	Lessens oxidative damage of cells,and prevents lung damage from pollutants; vital to the immune system	Deficiency is rare; anemia, nervous system degeneration	Wheat germ oil, unrefined vegetable oils, nuts, seeds, grains
Vitamin K (Phylloquinone)	Essential to blood clotting	Bleeding and easy bruising	Leafy green vegetables, tea

A healthy diet should include a variety of

foods to ensure that needs for essential nutrients are met. (credit: Keith Weller, USDA ARS)

Minerals and Their Function in the Human Body			
Mineral	**Function**	**Deficiencies Can Lead To**	**Sources**
*Calcium	Needed for muscle and neuron function; heart health; builds bone and supports synthesis and function of blood cells; nerve function	Osteoporosis, rickets, muscle spasms, impaired growth	Milk, yogurt, fish, green leafy vegetables, legumes
*Chlorine	Needed for production of hydrochloric acid (HCl) in the stomach and nerve function; osmotic balance	Muscle cramps, mood disturbances, reduced appetite	Table salt

Minerals and Their Function in the Human Body			
Mineral	**Function**	**Deficiencies Can Lead To**	**Sources**
Copper (trace amounts)	Required component of many redox enzymes, including cytochrome c oxidase; cofactor for hemoglobin synthesis	Copper deficiency is rare	Liver, oysters, cocoa, chocolate, sesame, nuts
Iodine	Required for the synthesis of thyroid hormones	Goiter	Seafood, iodized salt, dairy products
Iron	Required for many proteins and enzymes, notably hemoglobin, to prevent anemia	Anemia, which causes poor concentration, fatigue, and poor immune function	Red meat, leafy green vegetables, fish (tuna, salmon), eggs, dried fruits, beans, whole grains

Minerals and Their Function in the Human Body			
Mineral	**Function**	**Deficiencies Can Lead To**	**Sources**
*Magnesium	Required co-factor for ATP formation; bone formation; normal membrane functions; muscle function	Mood disturbances, muscle spasms	Whole grains, leafy green vegetables
Manganese (trace amounts)	A cofactor in enzyme functions; trace amounts are required	Manganese deficiency is rare	Common in most foods
Molybdenum (trace amounts)	Acts as a cofactor for three essential enzymes in humans: sulfite oxidase, xanthine oxidase, and aldehyde oxidase	Molybdenum deficiency is rare	
*Phosphorus	A component of bones and teeth; helps regulate acid-base balance; nucleotide synthesis	Weakness, bone abnormalities, calcium loss	Milk, hard cheese, whole grains, meats

Minerals and Their Function in the Human Body			
Mineral	**Function**	**Deficiencies Can Lead To**	**Sources**
*Potassium	Vital for muscles, heart, and nerve function	Cardiac rhythm disturbance, muscle weakness	Legumes, potato skin, tomatoes, bananas
Selenium (trace amounts)	A cofactor essential to activity of antioxidant enzymes like glutathione peroxidase; trace amounts are required	Selenium deficiency is rare	Common in most foods
*Sodium	Systemic electrolyte required for many functions; acid-base balance; water balance; nerve function	Muscle cramps, fatigue, reduced appetite	Table salt
Zinc (trace amounts)	Required for several enzymes such as carboxypeptidase, liver alcohol dehydrogenase, and carbonic anhydrase	Anemia, poor wound healing, can lead to short stature	Common in most foods

Minerals and Their Function in the Human Body			
Mineral	**Function**	**Deficiencies Can Lead To**	**Sources**
*Greater than 200mg/day required			

Essential Amino Acids	
Amino acids that must be consumed	**Amino acids anabolized by the body**
isoleucine	alanine
leucine	selenocysteine
lysine	aspartate
methionine	cysteine
phenylalanine	glutamate
tryptophan	glycine
valine	proline
histidine*	serine
threonine	tyrosine
arginine*	asparagine

Essential Amino Acids	
Amino acids that must be consumed	**Amino acids anabolized by the body**
*The human body can synthesize histidine and arginine, but not in the quantities required, especially for growing children.	

Food Energy and ATP

Animals need food to obtain energy and maintain homeostasis. Homeostasis is the ability of a system to maintain a stable internal environment even in the face of external changes to the environment. For example, the normal body temperature of humans is 37°C (98.6°F). Humans maintain this temperature even when the external temperature is hot or cold. It takes energy to maintain this body temperature, and animals obtain this energy from food.

The primary source of energy for animals is carbohydrates, mainly glucose. Glucose is called the body’s fuel. The digestible carbohydrates in an animal’s diet are converted to glucose molecules through a series of catabolic chemical reactions.

Adenosine triphosphate, or ATP, is the primary energy currency in cells; ATP stores energy in phosphate ester bonds. ATP releases energy when the phosphodiester bonds are broken and ATP is converted to ADP and a phosphate group. ATP is produced by the oxidative reactions in the cytoplasm and mitochondrion of the cell, where carbohydrates, proteins, and fats undergo a series of metabolic reactions collectively called cellular respiration. For example, glycolysis is a series of reactions in which glucose is converted to pyruvic acid and some of its chemical potential energy is transferred to NADH and ATP.

ATP is required for all cellular functions. It is used to build the organic molecules that are required for cells and tissues; it provides energy for muscle contraction and for the transmission of electrical signals in the nervous system. When the amount of ATP is available in excess of the body’s requirements, the liver uses the excess ATP and excess glucose to produce

molecules called glycogen. Glycogen is a polymeric form of glucose and is stored in the liver and skeletal muscle cells. When blood sugar drops, the liver releases glucose from stores of glycogen. Skeletal muscle converts glycogen to glucose during intense exercise. The process of converting glucose and excess ATP to glycogen and the storage of excess energy is an evolutionarily important step in helping animals deal with mobility, food shortages, and famine.

Note:

Everyday Connection

Obesity

Obesity is a major health concern in the United States, and there is a growing focus on reducing obesity and the diseases it may lead to, such as type-2 diabetes, cancers of the colon and breast, and cardiovascular disease. How does the food consumed contribute to obesity?

Fatty foods are calorie-dense, meaning that they have more calories per unit mass than carbohydrates or proteins. One gram of carbohydrates has four calories, one gram of protein has four calories, and one gram of fat has nine calories. Animals tend to seek lipid-rich food for their higher energy content. The signals of hunger ("time to eat") and satiety ("time to stop eating") are controlled in the hypothalamus region of the brain. Foods that are rich in fatty acids tend to promote satiety more than foods that are rich only in carbohydrates.

Excess carbohydrate and ATP are used by the liver to synthesize glycogen. The pyruvate produced during glycolysis is used to synthesize fatty acids. When there is more glucose in the body than required, the resulting excess pyruvate is converted into molecules that eventually result in the synthesis of fatty acids within the body. These fatty acids are stored in adipose cells—the fat cells in the mammalian body whose primary role is to store fat for later use.

It is important to note that some animals benefit from obesity. Polar bears and seals need body fat for insulation and to keep them from losing body heat during Arctic winters. When food is scarce, stored body fat provides energy for maintaining homeostasis. Fats prevent famine in mammals, allowing them to access energy when food is not available on a daily basis; fats are stored when a large kill is made or lots of food is available.

Section Summary

Animal diet should be balanced and meet the needs of the body. Carbohydrates, proteins, and fats are the primary components of food. Some essential nutrients are required for cellular function but cannot be produced by the animal body. These include vitamins, minerals, some fatty acids, and some amino acids. Food intake in more than necessary amounts is stored as glycogen in the liver and muscle cells, and in fat cells. Excess adipose storage can lead to obesity and serious health problems. ATP is the energy currency of the cell and is obtained from the metabolic pathways. Excess carbohydrates and energy are stored as glycogen in the body.

Review Questions

Exercise:

Problem: Which of the following statements is not true?

a. Essential nutrients can be synthesized by the body.
b. Vitamins are required in small quantities for bodily function.
c. Some amino acids can be synthesized by the body, while others need to be obtained from diet.
d. Vitamins come in two categories: fat-soluble and water-soluble.

Solution:

A

Exercise:

Problem: Which of the following is a water-soluble vitamin?

a. vitamin A
b. vitamin E
c. vitamin K
d. vitamin C

Solution:

D

Exercise:

Problem: What is the primary fuel for the body?

a. carbohydrates
b. lipids
c. protein
d. glycogen

Solution:

A

Exercise:

Problem: Excess glucose is stored as ________.

a. fat
b. glucagon
c. glycogen
d. it is not stored in the body

Solution:

C

Free Response

Exercise:

Problem: What are essential nutrients?

Solution:

Essential nutrients are those nutrients that must be obtained from the diet because they cannot be produced by the body. Vitamins and minerals are examples of essential nutrients.

Exercise:

Problem: What is the role of minerals in maintaining good health?

Solution:

Minerals—such as potassium, sodium, and calcium—are required for the functioning of many cellular processes, including muscle contraction and nerve conduction. While minerals are required in trace amounts, not having minerals in the diet can be potentially harmful.

Exercise:

Problem: Discuss why obesity is a growing epidemic.

Solution:

In the United States, obesity, particularly childhood obesity, is a growing concern. Some of the contributors to this situation include sedentary lifestyles and consuming more processed foods and less fruits and vegetables. As a result, even young children who are obese can face health concerns.

Exercise:

Problem:

There are several nations where malnourishment is a common occurrence. What may be some of the health challenges posed by malnutrition?

Solution:

Malnutrition, often in the form of not getting enough calories or not enough of the essential nutrients, can have severe consequences. Many malnourished children have vision and dental problems, and over the years may develop many serious health problems.

Glossary

essential nutrient
: nutrient that cannot be synthesized by the body; it must be obtained from food

mineral
: inorganic, elemental molecule that carries out important roles in the body

vitamin
: organic substance necessary in small amounts to sustain life

Digestive System Processes
By the end of this section, you will be able to:

- Describe the process of digestion
- Detail the steps involved in digestion and absorption
- Define elimination
- Explain the role of both the small and large intestines in absorption

Obtaining nutrition and energy from food is a multi-step process. For true animals, the first step is ingestion, the act of taking in food. This is followed by digestion, absorption, and elimination. In the following sections, each of these steps will be discussed in detail.

Ingestion

The large molecules found in intact food cannot pass through the cell membranes. Food needs to be broken into smaller particles so that animals can harness the nutrients and organic molecules. The first step in this process is **ingestion**. Ingestion is the process of taking in food through the mouth. In vertebrates, the teeth, saliva, and tongue play important roles in mastication (preparing the food into bolus). While the food is being mechanically broken down, the enzymes in saliva begin to chemically process the food as well. The combined action of these processes modifies the food from large particles to a soft mass that can be swallowed and can travel the length of the esophagus.

Digestion and Absorption

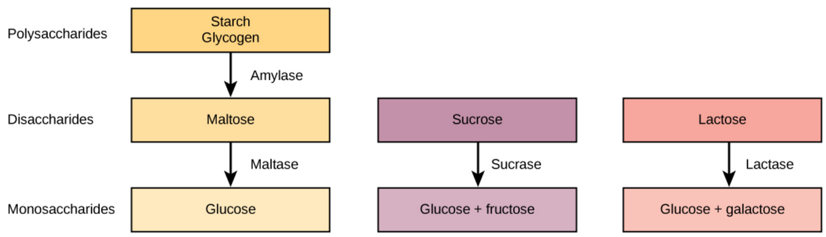
Digestion is the mechanical and chemical break down of food into small organic fragments. It is important to break down macromolecules into smaller fragments that are of suitable size for absorption across the digestive epithelium. Large, complex molecules of proteins, polysaccharides, and lipids must be reduced to simpler particles such as simple sugar before they can be absorbed by the digestive epithelial cells. Different organs play specific roles in the digestive process. The animal diet needs carbohydrates, protein, and fat, as well as vitamins and inorganic components for nutritional balance. How each of these components is digested is discussed in the following sections.

Carbohydrates

The digestion of carbohydrates begins in the mouth. The salivary enzyme amylase begins the breakdown of food starches into maltose, a disaccharide. As the bolus of food travels through the esophagus to the stomach, no significant digestion of carbohydrates takes place. The esophagus produces no digestive enzymes but does produce mucous for lubrication. The acidic environment in the stomach stops the action of the amylase enzyme.

The next step of carbohydrate digestion takes place in the duodenum. Recall that the chyme from the stomach enters the duodenum and mixes with the digestive secretion from the pancreas, liver, and gallbladder. Pancreatic juices also contain amylase, which continues the breakdown of starch and glycogen into maltose, a disaccharide. The disaccharides are broken down into

monosaccharides by enzymes called **maltases**, **sucrases**, and **lactases**, which are also present in the brush border of the small intestinal wall. Maltase breaks down maltose into glucose. Other disaccharides, such as sucrose and lactose are broken down by sucrase and lactase, respectively. Sucrase breaks down sucrose (or "table sugar") into glucose and fructose, and lactase breaks down lactose (or "milk sugar") into glucose and galactose. The monosaccharides (glucose) thus produced are absorbed and then can be used in metabolic pathways to harness energy. The monosaccharides are transported across the intestinal epithelium into the bloodstream to be transported to the different cells in the body. The steps in carbohydrate digestion are summarized in [link] and [link].



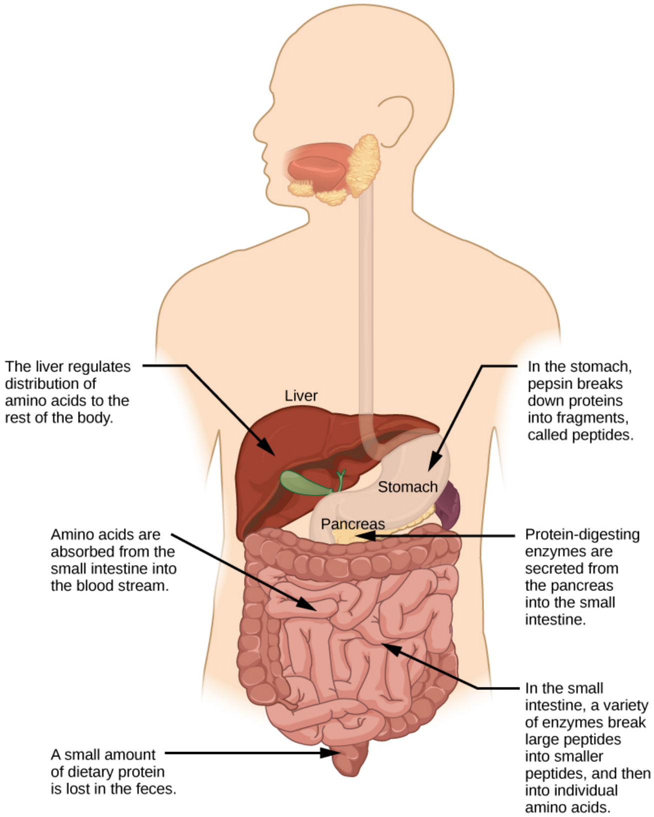
Digestion of carbohydrates is performed by several enzymes. Starch and glycogen are broken down into glucose by amylase and maltase. Sucrose (table sugar) and lactose (milk sugar) are broken down by sucrase and lactase, respectively.

Digestion of Carbohydrates				
Enzyme	**Produced By**	**Site of Action**	**Substrate Acting On**	**End Products**
Salivary amylase	Salivary glands	Mouth	Polysaccharides (Starch)	Disaccharides (maltose), oligosaccharides
Pancreatic amylase	Pancreas	Small intestine	Polysaccharides (starch)	Disaccharides (maltose), monosaccharides

Digestion of Carbohydrates				
Enzyme	**Produced By**	**Site of Action**	**Substrate Acting On**	**End Products**
Oligosaccharidases	Lining of the intestine; brush border membrane	Small intestine	Disaccharides	Monosaccharides (e.g., glucose, fructose, galactose)

Protein

A large part of protein digestion takes place in the stomach. The enzyme pepsin plays an important role in the digestion of proteins by breaking down the intact protein to peptides, which are short chains of four to nine amino acids. In the duodenum, other enzymes—**trypsin**, **elastase**, and **chymotrypsin**—act on the peptides reducing them to smaller peptides. Trypsin elastase, carboxypeptidase, and chymotrypsin are produced by the pancreas and released into the duodenum where they act on the chyme. Further breakdown of peptides to single amino acids is aided by enzymes called peptidases (those that break down peptides). Specifically, **carboxypeptidase**, **dipeptidase**, and **aminopeptidase** play important roles in reducing the peptides to free amino acids. The amino acids are absorbed into the bloodstream through the small intestines. The steps in protein digestion are summarized in [link] and [link].



Protein digestion is a multistep process that begins in the stomach and continues through the intestines.

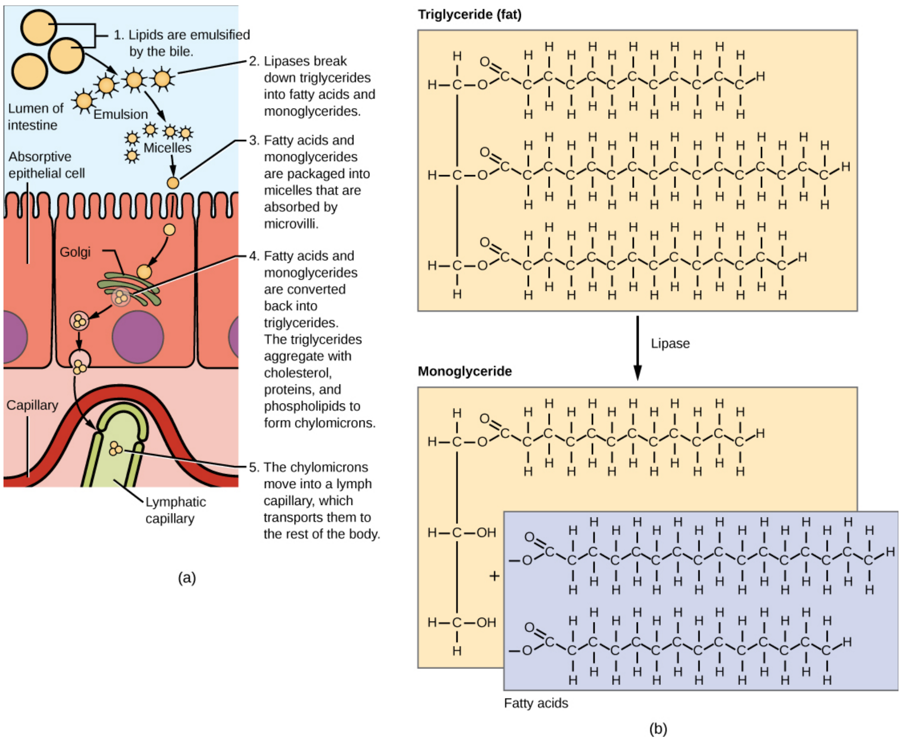
Digestion of Protein				
Enzyme	**Produced By**	**Site of Action**	**Substrate Acting On**	**End Products**
Pepsin	Stomach chief cells	Stomach	Proteins	Peptides
Trypsin Elastase Chymotrypsin	Pancreas	Small intestine	Proteins	Peptides

Digestion of Protein				
Enzyme	**Produced By**	**Site of Action**	**Substrate Acting On**	**End Products**
Carboxypeptidase	Pancreas	Small intestine	Peptides	Amino acids and peptides
Aminopeptidase Dipeptidase	Lining of intestine	Small intestine	Peptides	Amino acids

Lipids

Lipid digestion begins in the stomach with the aid of lingual lipase and gastric lipase. However, the bulk of lipid digestion occurs in the small intestine due to pancreatic lipase. When chyme enters the duodenum, the hormonal responses trigger the release of bile, which is produced in the liver and stored in the gallbladder. Bile aids in the digestion of lipids, primarily triglycerides by emulsification. Emulsification is a process in which large lipid globules are broken down into several small lipid globules. These small globules are more widely distributed in the chyme rather than forming large aggregates. Lipids are hydrophobic substances: in the presence of water, they will aggregate to form globules to minimize exposure to water. Bile contains bile salts, which are amphipathic, meaning they contain hydrophobic and hydrophilic parts. Thus, the bile salts hydrophilic side can interface with water on one side and the hydrophobic side interfaces with lipids on the other. By doing so, bile salts emulsify large lipid globules into small lipid globules.

Why is emulsification important for digestion of lipids? Pancreatic juices contain enzymes called lipases (enzymes that break down lipids). If the lipid in the chyme aggregates into large globules, very little surface area of the lipids is available for the lipases to act on, leaving lipid digestion incomplete. By forming an emulsion, bile salts increase the available surface area of the lipids many fold. The pancreatic lipases can then act on the lipids more efficiently and digest them, as detailed in [link]. Lipases break down the lipids into fatty acids and glycerides. These molecules can pass through the plasma membrane of the cell and enter the epithelial cells of the intestinal lining. The bile salts surround long-chain fatty acids and monoglycerides forming tiny spheres called micelles. The micelles move into the brush border of the small intestine absorptive cells where the long-chain fatty acids and monoglycerides diffuse out of the micelles into the absorptive cells leaving the micelles behind in the chyme. The long-chain fatty acids and monoglycerides recombine in the absorptive cells to form triglycerides, which aggregate into globules and become coated with proteins. These large spheres are called **chylomicrons**. Chylomicrons contain triglycerides, cholesterol, and other lipids and have proteins on their surface. The surface is also composed of the hydrophilic phosphate "heads" of phospholipids. Together, they enable the chylomicron to move in an aqueous environment without exposing the lipids to water. Chylomicrons leave the absorptive cells via exocytosis. Chylomicrons enter the lymphatic vessels, and then enter the blood in the subclavian vein.



Lipids are digested and absorbed in the small intestine.

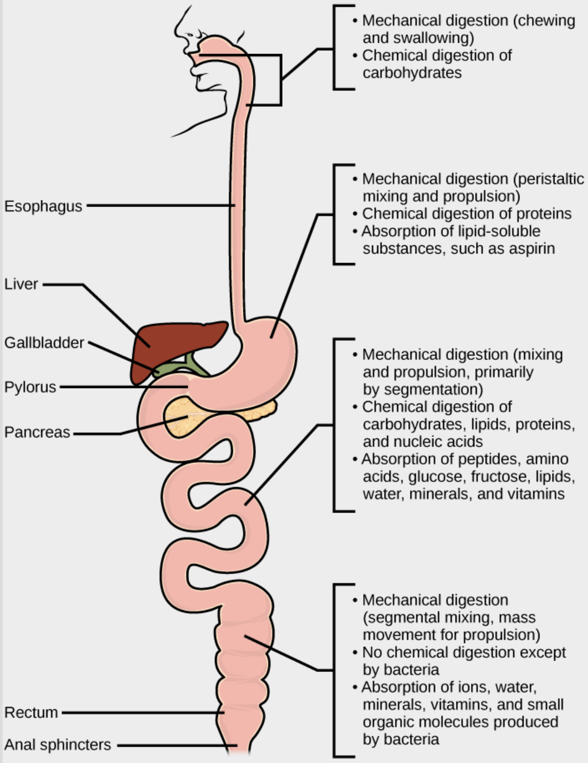
Vitamins

Vitamins can be either water-soluble or lipid-soluble. Fat soluble vitamins are absorbed in the same manner as lipids. It is important to consume some amount of dietary lipid to aid the absorption of lipid-soluble vitamins. Water-soluble vitamins can be directly absorbed into the bloodstream from the intestine.

Note:
Link to Learning

This website has an overview of the digestion of protein, fat, and carbohydrates.

Note:
Art Connection



Mechanical and chemical digestion of food takes place in many steps, beginning in the mouth and ending in the rectum.

Which of the following statements about digestive processes is true?

a. Amylase, maltase, and lactase in the mouth digest carbohydrates.
b. Trypsin and lipase in the stomach digest protein.
c. Bile emulsifies lipids in the small intestine.
d. No food is absorbed until the small intestine.

Elimination

The final step in digestion is the elimination of undigested food content and waste products. The undigested food material enters the colon, where most of the water is reabsorbed. Recall that the colon is also home to the microflora called “intestinal flora” that aid in the digestion process. The semi-solid waste is moved through the colon by peristaltic movements of the muscle and is stored in the rectum. As the rectum expands in response to storage of fecal matter, it triggers the neural signals required to set up the urge to eliminate. The solid waste is eliminated through the anus using peristaltic movements of the rectum.

Common Problems with Elimination

Diarrhea and constipation are some of the most common health concerns that affect digestion. Constipation is a condition where the feces are hardened because of excess water removal in the colon. In contrast, if enough water is not removed from the feces, it results in diarrhea. Many bacteria, including the ones that cause cholera, affect the proteins involved in water reabsorption in the colon and result in excessive diarrhea.

Emesis

Emesis, or vomiting, is elimination of food by forceful expulsion through the mouth. It is often in response to an irritant that affects the digestive tract, including but not limited to viruses, bacteria, emotions, sights, and food poisoning. This forceful expulsion of the food is due to the strong contractions produced by the stomach muscles. The process of emesis is regulated by the medulla.

Section Summary

Digestion begins with ingestion, where the food is taken in the mouth. Digestion and absorption take place in a series of steps with special enzymes playing important roles in digesting carbohydrates, proteins, and lipids. Elimination describes removal of undigested food contents and waste products from the body. While most absorption occurs in the small intestines, the large intestine is responsible for the final removal of water that remains after the absorptive process of the small intestines. The cells that line the large intestine absorb some vitamins as well as any leftover salts and water. The large intestine (colon) is also where feces is formed.

Art Connections

Exercise:

Problem: [link] Which of the following statements about digestive processes is true?

a. Amylase, maltase and lactase in the mouth digest carbohydrates.
b. Trypsin and lipase in the stomach digest protein.
c. Bile emulsifies lipids in the small intestine.
d. No food is absorbed until the small intestine.

Solution:

[link] C

Review Questions

Exercise:

Problem: Where does the majority of protein digestion take place?

a. stomach
b. duodenum
c. mouth
d. jejunum

Solution:

A

Exercise:

Problem: Lipases are enzymes that break down _________.

a. disaccharides
b. lipids
c. proteins
d. cellulose

Solution:

B

Free Response

Exercise:

Problem: Explain why some dietary lipid is a necessary part of a balanced diet.

Solution:

Lipids add flavor to food and promote a sense of satiety or fullness. Fatty foods are sources of high energy; one gram of lipid contains nine calories. Lipids are also required in the diet to aid the absorption of lipid-soluble vitamins and for the production of lipid-soluble hormones.

Glossary

aminopeptidase
: protease that breaks down peptides to single amino acids; secreted by the brush border of small intestine

carboxypeptidase

protease that breaks down peptides to single amino acids; secreted by the brush border of the small intestine

chylomicron
: small lipid globule

chymotrypsin
: pancreatic protease

digestion
: mechanical and chemical break down of food into small organic fragments

dipeptidase
: protease that breaks down peptides to single amino acids; secreted by the brush border of small intestine

elastase
: pancreatic protease

ingestion
: act of taking in food

lactase
: enzyme that breaks down lactose into glucose and galactose

maltase
: enzyme that breaks down maltose into glucose

sucrase
: enzyme that breaks down sucrose into glucose and fructose

trypsin
: pancreatic protease that breaks down protein

Digestive System Regulation
By the end of this section, you will be able to:

- Discuss the role of neural regulation in digestive processes
- Explain how hormones regulate digestion

The brain is the control center for the sensation of hunger and satiety. The functions of the digestive system are regulated through neural and hormonal responses.

Neural Responses to Food

In reaction to the smell, sight, or thought of food, like that shown in [link], the first response is that of salivation. The salivary glands secrete more saliva in response to stimulation by the autonomic nervous system triggered by food in preparation for digestion. Simultaneously, the stomach begins to produce hydrochloric acid to digest the food. Recall that the peristaltic movements of the esophagus and other organs of the digestive tract are under the control of the brain. The brain prepares these muscles for movement as well. When the stomach is full, the part of the brain that detects satiety signals fullness. There are three overlapping phases of gastric control—the cephalic phase, the gastric phase, and the intestinal phase—each requires many enzymes and is under neural control as well.

Seeing a plate of food triggers the secretion of saliva in the mouth and the production of HCL in the stomach. (credit: Kelly Bailey)

Digestive Phases

The response to food begins even before food enters the mouth. The first phase of ingestion, called the **cephalic phase**, is controlled by the neural response to the stimulus provided by food. All aspects—such as sight, sense, and smell—trigger the neural responses resulting in salivation and secretion of gastric juices. The gastric and salivary secretion in the cephalic phase can also take place due to the thought of food. Right now, if you think about a piece of chocolate or a crispy potato chip, the increase in salivation is a cephalic phase response to the thought. The central nervous system prepares the stomach to receive food.

The **gastric phase** begins once the food arrives in the stomach. It builds on the stimulation provided during the cephalic phase. Gastric acids and enzymes process the ingested materials. The gastric phase is stimulated by

(1) distension of the stomach, (2) a decrease in the pH of the gastric contents, and (3) the presence of undigested material. This phase consists of local, hormonal, and neural responses. These responses stimulate secretions and powerful contractions.

The **intestinal phase** begins when chyme enters the small intestine triggering digestive secretions. This phase controls the rate of gastric emptying. In addition to gastrin emptying, when chyme enters the small intestine, it triggers other hormonal and neural events that coordinate the activities of the intestinal tract, pancreas, liver, and gallbladder.

Hormonal Responses to Food

The **endocrine system** controls the response of the various glands in the body and the release of hormones at the appropriate times.

One of the important factors under hormonal control is the stomach acid environment. During the gastric phase, the hormone **gastrin** is secreted by G cells in the stomach in response to the presence of proteins. Gastrin stimulates the release of stomach acid, or hydrochloric acid (HCl) which aids in the digestion of the proteins. However, when the stomach is emptied, the acidic environment need not be maintained and a hormone called **somatostatin** stops the release of hydrochloric acid. This is controlled by a negative feedback mechanism.

In the duodenum, digestive secretions from the liver, pancreas, and gallbladder play an important role in digesting chyme during the intestinal phase. In order to neutralize the acidic chyme, a hormone called **secretin** stimulates the pancreas to produce alkaline bicarbonate solution and deliver it to the duodenum. Secretin acts in tandem with another hormone called **cholecystokinin** (CCK). Not only does CCK stimulate the pancreas to produce the requisite pancreatic juices, it also stimulates the gallbladder to release bile into the duodenum.

Note:

Link to Learning

Visit this website to learn more about the endocrine system. Review the text and watch the animation of how control is implemented in the endocrine system.

Another level of hormonal control occurs in response to the composition of food. Foods high in lipids take a long time to digest. A hormone called **gastric inhibitory peptide** is secreted by the small intestine to slow down the peristaltic movements of the intestine to allow fatty foods more time to be digested and absorbed.

Understanding the hormonal control of the digestive system is an important area of ongoing research. Scientists are exploring the role of each hormone in the digestive process and developing ways to target these hormones. Advances could lead to knowledge that may help to battle the obesity epidemic.

Section Summary

The brain and the endocrine system control digestive processes. The brain controls the responses of hunger and satiety. The endocrine system controls the release of hormones and enzymes required for digestion of food in the digestive tract.

Review Questions

Exercise:

Problem:

Which hormone controls the release of bile from the gallbladder

a. pepsin
b. amylase
c. CCK
d. gastrin

Solution:

C

Exercise:

Problem: Which hormone stops acid secretion in the stomach?

a. gastrin
b. somatostatin
c. gastric inhibitory peptide
d. CCK

Solution:

B

Free Response

Exercise:

Problem: Describe how hormones regulate digestion.

Solution:

Hormones control the different digestive enzymes that are secreted in the stomach and the intestine during the process of digestion and absorption. For example, the hormone gastrin stimulates stomach acid secretion in response to food intake. The hormone somatostatin stops the release of stomach acid.

Exercise:

Problem:

Describe one or more scenarios where loss of hormonal regulation of digestion can lead to diseases.

Solution:

There are many cases where loss of hormonal regulation can lead to illnesses. For example, the bilirubin produced by the breakdown of red blood cells is converted to bile by the liver. When there is malfunction of this process, there is excess bilirubin in the blood and bile levels are low. As a result, the body struggles with dealing with fatty food. This is why a patient suffering from jaundice is asked to eat a diet with almost zero fat.

Glossary

cephalic phase
: first phase of digestion, controlled by the neural response to the stimulus provided by food

cholecystokinin
: hormone that stimulates the contraction of the gallbladder to release bile

endocrine system
: system that controls the response of the various glands in the body and the release of hormones at the appropriate times

gastric inhibitory peptide

hormone secreted by the small intestine in the presence of fatty acids and sugars; it also inhibits acid production and peristalsis in order to slow down the rate at which food enters the small intestine

gastric phase
digestive phase beginning once food enters the stomach; gastric acids and enzymes process the ingested materials

gastrin
hormone which stimulates hydrochloric acid secretion in the stomach

intestinal phase
third digestive phase; begins when chyme enters the small intestine triggering digestive secretions and controlling the rate of gastric emptying

secretin
hormone which stimulates sodium bicarbonate secretion in the small intestine

somatostatin
hormone released to stop acid secretion when the stomach is empty

Overview of the Circulatory System
By the end of this section, you will be able to:

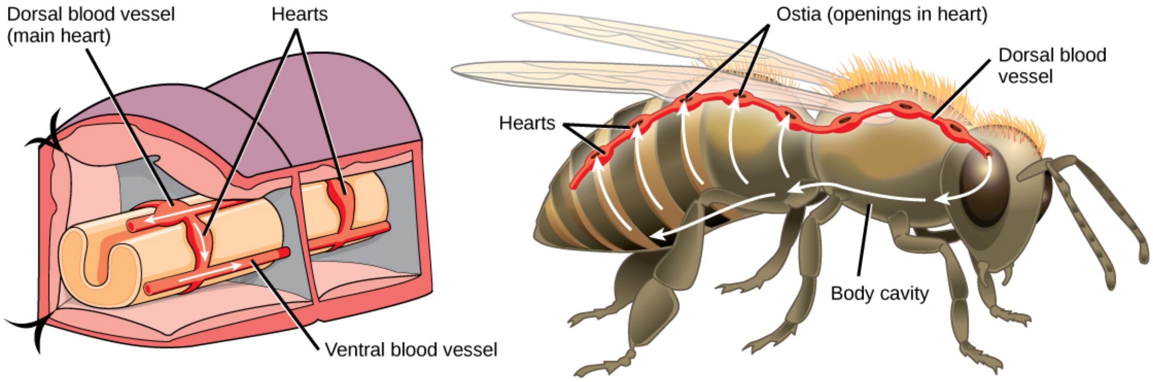
- Describe an open and closed circulatory system
- Describe interstitial fluid and hemolymph
- Compare and contrast the organization and evolution of the vertebrate circulatory system.

In all animals, except a few simple types, the circulatory system is used to transport nutrients and gases through the body. Simple diffusion allows some water, nutrient, waste, and gas exchange into primitive animals that are only a few cell layers thick; however, bulk flow is the only method by which the entire body of larger more complex organisms is accessed.

Circulatory System Architecture

The circulatory system is effectively a network of cylindrical vessels: the arteries, veins, and capillaries that emanate from a pump, the heart. In all vertebrate organisms, as well as some invertebrates, this is a closed-loop system, in which the blood is not free in a cavity. In a **closed circulatory system**, blood is contained inside blood vessels and circulates **unidirectionally** from the heart around the systemic circulatory route, then returns to the heart again, as illustrated in [link]**a**. As opposed to a closed system, arthropods—including insects, crustaceans, and most mollusks—have an open circulatory system, as illustrated in [link]**b**. In an **open circulatory system**, the blood is not enclosed in the blood vessels but is pumped into a cavity called a **hemocoel** and is called **hemolymph** because the blood mixes with the **interstitial fluid**. As the heart beats and the animal moves, the hemolymph circulates around the organs within the body cavity and then reenters the hearts through openings called **ostia**. This movement allows for gas and nutrient exchange. An open circulatory system does not use as much energy as a closed system to operate or to maintain; however, there is a trade-off with the amount of blood that can be moved to metabolically active organs and tissues that require high levels of oxygen. In fact, one reason that insects with wing spans of up to two feet wide (70 cm) are not around today is probably because they were outcompeted by the arrival of birds 150 million years ago. Birds, having a

closed circulatory system, are thought to have moved more agilely, allowing them to get food faster and possibly to prey on the insects.

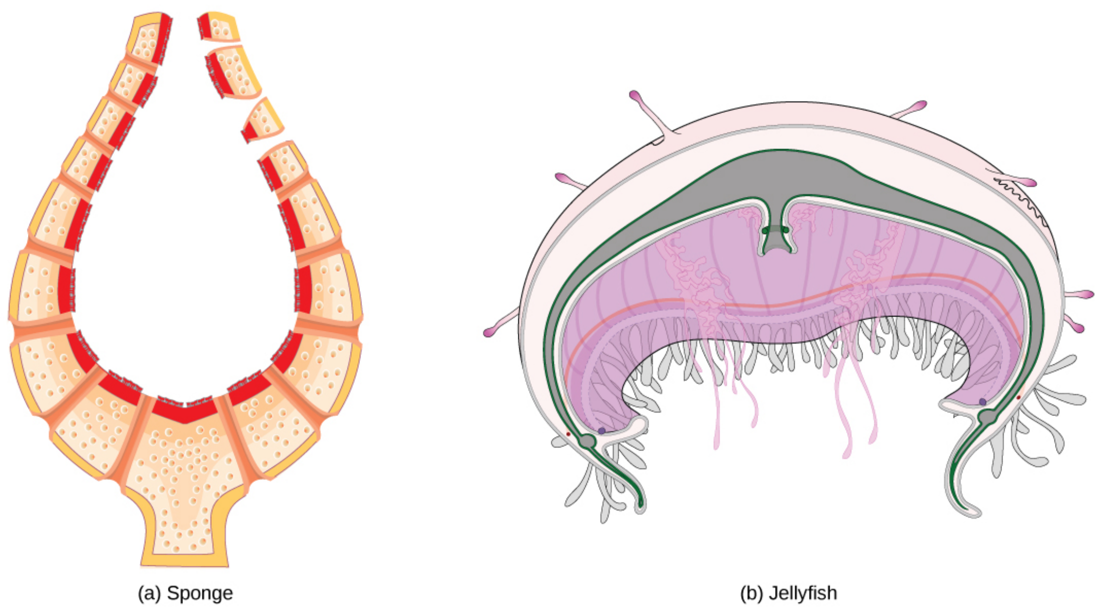


In (a) closed circulatory systems, the heart pumps blood through vessels that are separate from the interstitial fluid of the body. Most vertebrates and some invertebrates, like this annelid earthworm, have a closed circulatory system. In (b) open circulatory systems, a fluid called hemolymph is pumped through a blood vessel that empties into the body cavity. Hemolymph returns to the blood vessel through openings called ostia. Arthropods like this bee and most mollusks have open circulatory systems.

Circulatory System Variation in Animals

The circulatory system varies from simple systems in invertebrates to more complex systems in vertebrates. The simplest animals, such as the sponges (Porifera) and rotifers (Rotifera), do not need a circulatory system because diffusion allows adequate exchange of water, nutrients, and waste, as well as dissolved gases, as shown in [link]**a**. Organisms that are more complex but still only have two layers of cells in their body plan, such as jellies (Cnidaria) and comb jellies (Ctenophora) also use diffusion through their epidermis and internally through the gastrovascular compartment. Both

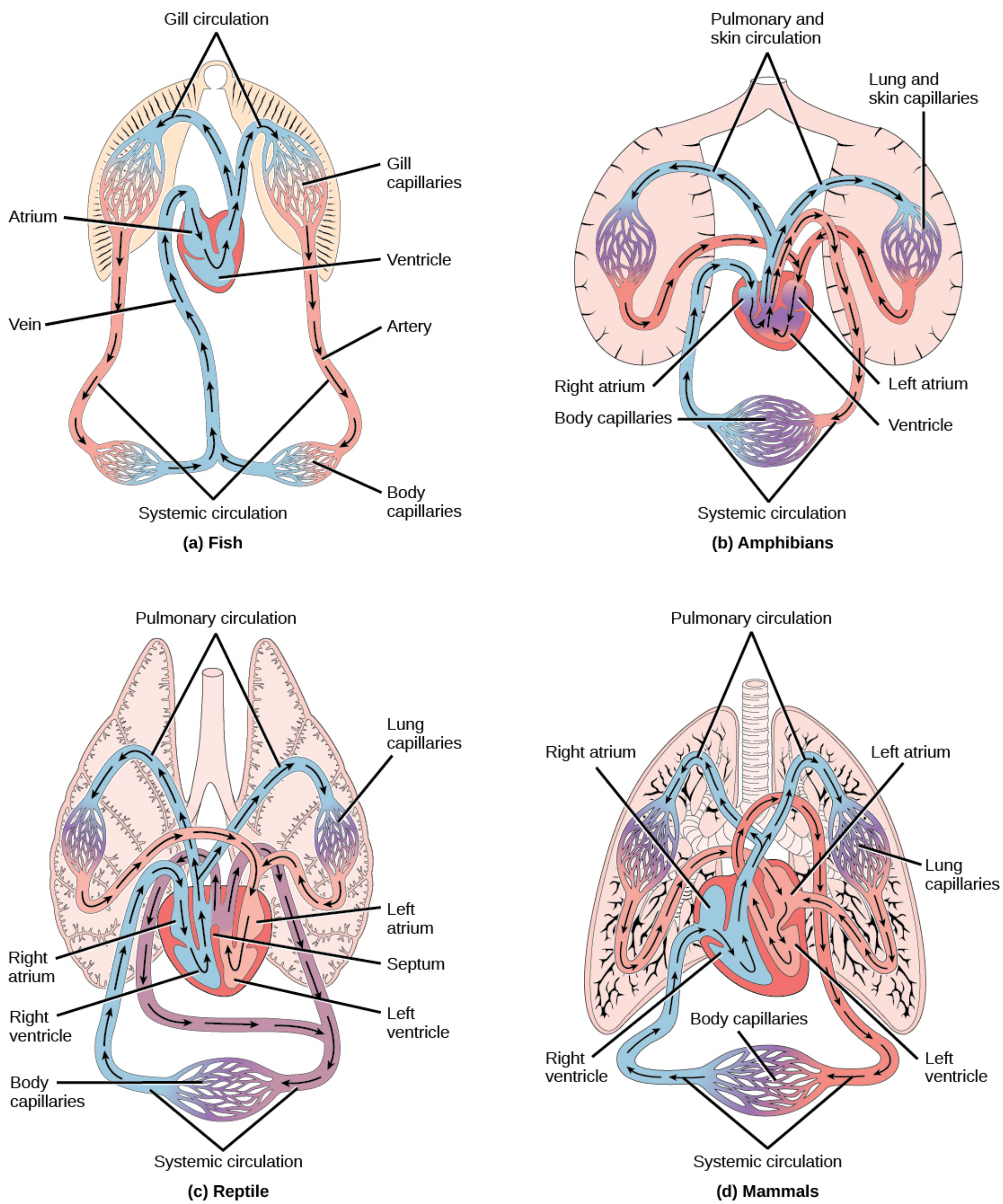
their internal and external tissues are bathed in an aqueous environment and exchange fluids by diffusion on both sides, as illustrated in [link]**b**. Exchange of fluids is assisted by the pulsing of the jellyfish body.



Simple animals consisting of a single cell layer such as the (a) sponge or only a few cell layers such as the (b) jellyfish do not have a circulatory system. Instead, gases, nutrients, and wastes are exchanged by diffusion.

For more complex organisms, diffusion is not efficient for cycling gases, nutrients, and waste effectively through the body; therefore, more complex circulatory systems evolved. Most arthropods and many mollusks have open circulatory systems. In an open system, an elongated beating heart pushes the hemolymph through the body and muscle contractions help to move fluids. The larger more complex crustaceans, including lobsters, have developed arterial-like vessels to push blood through their bodies, and the most active mollusks, such as squids, have evolved a closed circulatory system and are able to move rapidly to catch prey. Closed circulatory systems are a characteristic of vertebrates; however, there are significant

differences in the structure of the heart and the circulation of blood between the different vertebrate groups due to adaptation during evolution and associated differences in anatomy. [link] illustrates the basic circulatory systems of some vertebrates: fish, amphibians, reptiles, and mammals.



(a) Fish have the simplest circulatory systems of the vertebrates: blood flows unidirectionally from the two-chambered heart through the gills and then the rest of the body. (b) Amphibians have two circulatory routes: one for oxygenation of the blood through the lungs and skin, and the other to take oxygen to the rest of the body. The blood is pumped from a three-chambered heart with two atria and a single ventricle. (c) Reptiles also have two circulatory routes; however, blood is only oxygenated through the lungs. The heart is three chambered, but the ventricles are partially separated so some mixing of oxygenated and deoxygenated blood occurs except in crocodilians and birds. (d) Mammals and birds have the most efficient heart with four chambers that completely separate the oxygenated and deoxygenated blood; it pumps only oxygenated blood through the body and deoxygenated blood to the lungs.

As illustrated in [link]**a** Fish have a single circuit for blood flow and a two-chambered heart that has only a single atrium and a single ventricle. The atrium collects blood that has returned from the body and the ventricle pumps the blood to the gills where gas exchange occurs and the blood is re-oxygenated; this is called **gill circulation**. The blood then continues through the rest of the body before arriving back at the atrium; this is called **systemic circulation**. This unidirectional flow of blood produces a gradient of oxygenated to deoxygenated blood around the fish's systemic circuit. The result is a limit in the amount of oxygen that can reach some of the organs and tissues of the body, reducing the overall metabolic capacity of fish.

In amphibians, reptiles, birds, and mammals, blood flow is directed in two circuits: one through the lungs and back to the heart, which is called **pulmonary circulation**, and the other throughout the rest of the body and its organs including the brain (systemic circulation). In amphibians, gas exchange also occurs through the skin during pulmonary circulation and is referred to as **pulmocutaneous circulation**.

As shown in [link]**b**, amphibians have a three-chambered heart that has two atria and one ventricle rather than the two-chambered heart of fish. The two **atria** (superior heart chambers) receive blood from the two different circuits (the lungs and the systems), and then there is some mixing of the blood in the heart's **ventricle** (inferior heart chamber), which reduces the efficiency of oxygenation. The advantage to this arrangement is that high pressure in the vessels pushes blood to the lungs and body. The mixing is mitigated by a ridge within the ventricle that diverts oxygen-rich blood through the systemic circulatory system and deoxygenated blood to the pulmocutaneous circuit. For this reason, amphibians are often described as having **double circulation**.

Most reptiles also have a three-chambered heart similar to the amphibian heart that directs blood to the pulmonary and systemic circuits, as shown in [link]**c**. The ventricle is divided more effectively by a partial septum, which results in less mixing of oxygenated and deoxygenated blood. Some reptiles (alligators and crocodiles) are the most primitive animals to exhibit a four-chambered heart. Crocodilians have a unique circulatory mechanism where the heart shunts blood from the lungs toward the stomach and other organs during long periods of submergence, for instance, while the animal waits for prey or stays underwater waiting for prey to rot. One adaptation includes two main arteries that leave the same part of the heart: one takes blood to the lungs and the other provides an alternate route to the stomach and other parts of the body. Two other adaptations include a hole in the heart between the two ventricles, called the foramen of Panizza, which allows blood to move from one side of the heart to the other, and specialized connective tissue that slows the blood flow to the lungs. Together these adaptations have made crocodiles and alligators one of the most evolutionarily successful animal groups on earth.

In mammals and birds, the heart is also divided into four chambers: two atria and two ventricles, as illustrated in [link]**d**. The oxygenated blood is separated from the deoxygenated blood, which improves the efficiency of double circulation and is probably required for the warm-blooded lifestyle of mammals and birds. The four-chambered heart of birds and mammals evolved independently from a three-chambered heart. The independent

evolution of the same or a similar biological trait is referred to as convergent evolution.

Section Summary

In most animals, the circulatory system is used to transport blood through the body. Some primitive animals use diffusion for the exchange of water, nutrients, and gases. However, complex organisms use the circulatory system to carry gases, nutrients, and waste through the body. Circulatory systems may be open (mixed with the interstitial fluid) or closed (separated from the interstitial fluid). Closed circulatory systems are a characteristic of vertebrates; however, there are significant differences in the structure of the heart and the circulation of blood between the different vertebrate groups due to adaptions during evolution and associated differences in anatomy. Fish have a two-chambered heart with unidirectional circulation. Amphibians have a three-chambered heart, which has some mixing of the blood, and they have double circulation. Most non-avian reptiles have a three-chambered heart, but have little mixing of the blood; they have double circulation. Mammals and birds have a four-chambered heart with no mixing of the blood and double circulation.

Review Questions

Exercise:

Problem:

Why are open circulatory systems advantageous to some animals?

a. They use less metabolic energy.
b. They help the animal move faster.
c. They do not need a heart.
d. They help large insects develop.

Solution:

A

Exercise:

Problem:

Some animals use diffusion instead of a circulatory system. Examples include:

a. birds and jellyfish
b. flatworms and arthropods
c. mollusks and jellyfish
d. None of the above

Solution:

D

Exercise:

Problem:

Blood flow that is directed through the lungs and back to the heart is called _________.

a. unidirectional circulation
b. gill circulation
c. pulmonary circulation
d. pulmocutaneous circulation

Solution:

C

Free Response

Exercise:

Problem: Describe a closed circulatory system.

Solution:

A closed circulatory system is a closed-loop system, in which blood is not free in a cavity. Blood is separate from the bodily interstitial fluid and contained within blood vessels. In this type of system, blood circulates unidirectionally from the heart around the systemic circulatory route, and then returns to the heart.

Exercise:

Problem: Describe systemic circulation.

Solution:

Systemic circulation flows through the systems of the body. The blood flows away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body; it then returns to the heart.

Glossary

atrium
: (plural: atria) chamber of the heart that receives blood from the veins and sends blood to the ventricles

closed circulatory system
: system in which the blood is separated from the bodily interstitial fluid and contained in blood vessels

double circulation
: flow of blood in two circuits: the pulmonary circuit through the lungs and the systemic circuit through the organs and body

gill circulation

circulatory system that is specific to animals with gills for gas exchange; the blood flows through the gills for oxygenation

hemocoel
: cavity into which blood is pumped in an open circulatory system

hemolymph
: mixture of blood and interstitial fluid that is found in insects and other arthropods as well as most mollusks

interstitial fluid
: fluid between cells

open circulatory system
: system in which the blood is mixed with interstitial fluid and directly covers the organs

ostium
: (plural: ostia) holes between blood vessels that allow the movement of hemolymph through the body of insects, arthropods, and mollusks with open circulatory systems

pulmocutaneous circulation
: circulatory system in amphibians; the flow of blood to the lungs and the moist skin for gas exchange

pulmonary circulation
: flow of blood away from the heart through the lungs where oxygenation occurs and then returns to the heart again

systemic circulation
: flow of blood away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body, and then the return of this blood to the heart

unidirectional circulation
: flow of blood in a single circuit; occurs in fish where the blood flows through the gills, then past the organs and the rest of the body, before

returning to the heart

ventricle
(heart) large inferior chamber of the heart that pumps blood into arteries

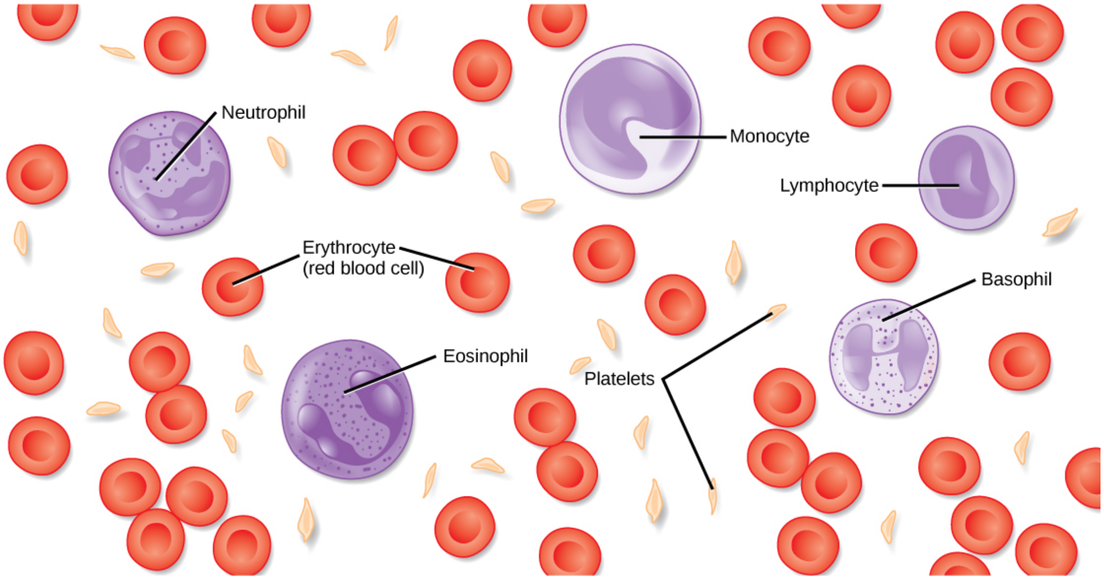
Components of the Blood
By the end of this section, you will be able to:

- List the basic components of the blood
- Compare red and white blood cells
- Describe blood plasma and serum

Hemoglobin is responsible for distributing oxygen, and to a lesser extent, carbon dioxide, throughout the circulatory systems of humans, vertebrates, and many invertebrates. The blood is more than the proteins, though. Blood is actually a term used to describe the liquid that moves through the vessels and includes **plasma** (the liquid portion, which contains water, proteins, salts, lipids, and glucose) and the cells (red and white cells) and cell fragments called **platelets**. Blood plasma is actually the dominant component of blood and contains the water, proteins, electrolytes, lipids, and glucose. The cells are responsible for carrying the gases (red cells) and immune the response (white). The platelets are responsible for blood clotting. Interstitial fluid that surrounds cells is separate from the blood, but in hemolymph, they are combined. In humans, cellular components make up approximately 45 percent of the blood and the liquid plasma 55 percent. Blood is 20 percent of a person's extracellular fluid and eight percent of weight.

The Role of Blood in the Body

Blood, like the human blood illustrated in [link] is important for regulation of the body's systems and homeostasis. Blood helps maintain homeostasis by stabilizing pH, temperature, osmotic pressure, and by eliminating excess heat. Blood supports growth by distributing nutrients and hormones, and by removing waste. Blood plays a protective role by transporting clotting factors and platelets to prevent blood loss and transporting the disease-fighting agents or **white blood cells** to sites of infection.



The cells and cellular components of human blood are shown. Red blood cells deliver oxygen to the cells and remove carbon dioxide. White blood cells—including neutrophils, monocytes, lymphocytes, eosinophils, and basophils—are involved in the immune response. Platelets form clots that prevent blood loss after injury.

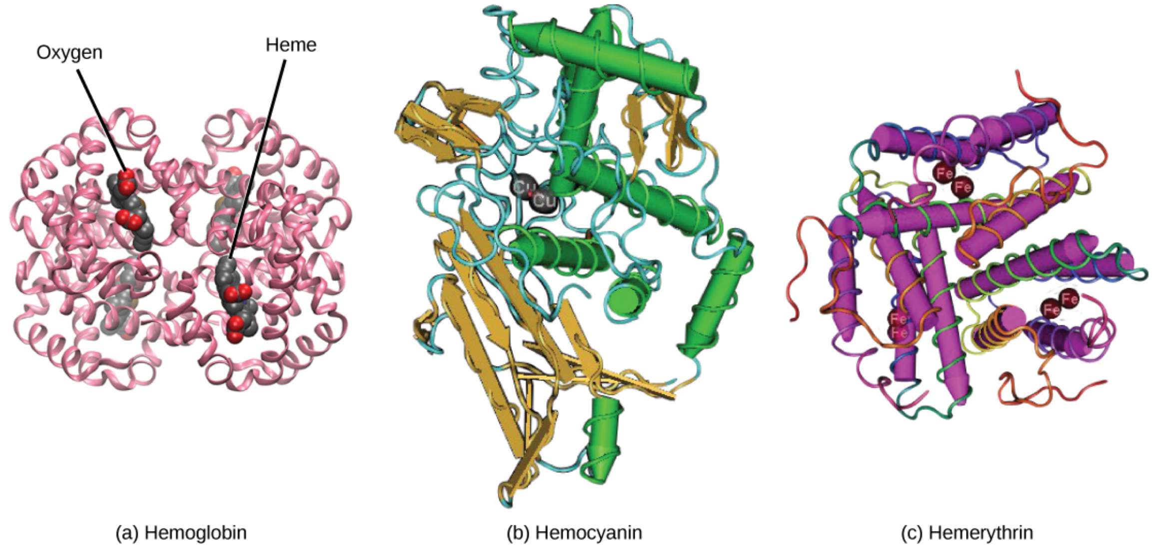
Red Blood Cells

Red blood cells, or erythrocytes (erythro- = “red”; -cyte = “cell”), are specialized cells that circulate through the body delivering oxygen to cells; they are formed from stem cells in the bone marrow. In mammals, red blood cells are small biconcave cells that at maturity do not contain a nucleus or mitochondria and are only 7–8 µm in size. In birds and non-avian reptiles, a nucleus is still maintained in red blood cells.

The red coloring of blood comes from the iron-containing protein hemoglobin, illustrated in [link]**a**. The principal job of this protein is to carry oxygen, but it also transports carbon dioxide as well. Hemoglobin is packed into red blood cells at a rate of about 250 million molecules of hemoglobin per cell. Each hemoglobin molecule binds four oxygen

molecules so that each red blood cell carries one billion molecules of oxygen. There are approximately 25 trillion red blood cells in the five liters of blood in the human body, which could carry up to 25 sextillion (25×10^{21}) molecules of oxygen in the body at any time. In mammals, the lack of organelles in erythrocytes leaves more room for the hemoglobin molecules, and the lack of mitochondria also prevents use of the oxygen for metabolic respiration. Only mammals have anucleated red blood cells, and some mammals (camels, for instance) even have nucleated red blood cells. The advantage of nucleated red blood cells is that these cells can undergo mitosis. Anucleated red blood cells metabolize anaerobically (without oxygen), making use of a primitive metabolic pathway to produce ATP and increase the efficiency of oxygen transport.

Not all organisms use hemoglobin as the method of oxygen transport. Invertebrates that utilize hemolymph rather than blood use different pigments to bind to the oxygen. These pigments use copper or iron to the oxygen. Invertebrates have a variety of other respiratory pigments. Hemocyanin, a blue-green, copper-containing protein, illustrated in [link]**b** is found in mollusks, crustaceans, and some of the arthropods. Chlorocruorin, a green-colored, iron-containing pigment is found in four families of polychaete tubeworms. Hemerythrin, a red, iron-containing protein is found in some polychaete worms and annelids and is illustrated in [link]**c**. Despite the name, hemerythrin does not contain a heme group and its oxygen-carrying capacity is poor compared to hemoglobin.



In most vertebrates, (a) hemoglobin delivers oxygen to the body and removes some carbon dioxide. Hemoglobin is composed of four protein subunits, two alpha chains and two beta chains, and a heme group that has iron associated with it. The iron reversibly associates with oxygen, and in so doing is oxidized from Fe^{2+} to Fe^{3+}. In most mollusks and some arthropods, (b) hemocyanin delivers oxygen. Unlike hemoglobin, hemolymph is not carried in blood cells, but floats free in the hemolymph. Copper instead of iron binds the oxygen, giving the hemolymph a blue-green color. In annelids, such as the earthworm, and some other invertebrates, (c) hemerythrin carries oxygen. Like hemoglobin, hemerythrin is carried in blood cells and has iron associated with it, but despite its name, hemerythrin does not contain heme.

The small size and large surface area of red blood cells allows for rapid diffusion of oxygen and carbon dioxide across the plasma membrane. In the lungs, carbon dioxide is released and oxygen is taken in by the blood. In the tissues, oxygen is released from the blood and carbon dioxide is bound for transport back to the lungs. Studies have found that hemoglobin also binds nitrous oxide (NO). NO is a vasodilator that relaxes the blood vessels and capillaries and may help with gas exchange and the passage of red blood cells through narrow vessels. Nitroglycerin, a heart medication for angina

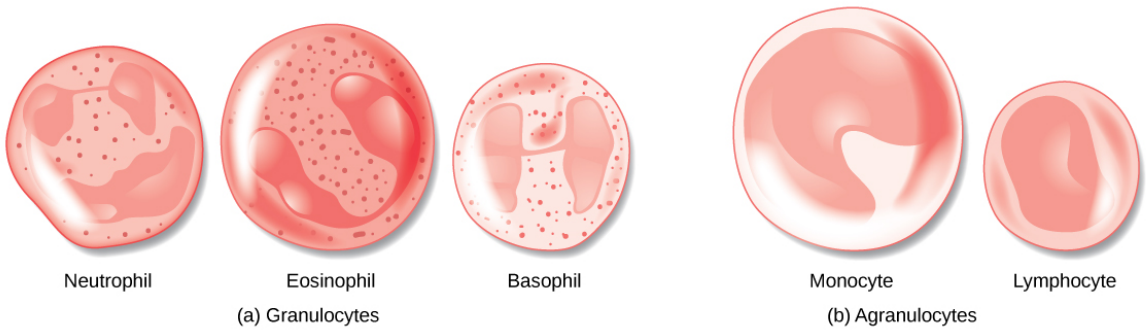
and heart attacks, is converted to NO to help relax the blood vessels and increase oxygen flow through the body.

A characteristic of red blood cells is their glycolipid and glycoprotein coating; these are lipids and proteins that have carbohydrate molecules attached. In humans, the surface glycoproteins and glycolipids on red blood cells vary between individuals, producing the different blood types, such as A, B, and O. Red blood cells have an average life span of 120 days, at which time they are broken down and recycled in the liver and spleen by phagocytic macrophages, a type of white blood cell.

White Blood Cells

White blood cells, also called leukocytes (leuko = white), make up approximately one percent by volume of the cells in blood. The role of white blood cells is very different than that of red blood cells: they are primarily involved in the immune response to identify and target pathogens, such as invading bacteria, viruses, and other foreign organisms. White blood cells are formed continually; some only live for hours or days, but some live for years.

The morphology of white blood cells differs significantly from red blood cells. They have nuclei and do not contain hemoglobin. The different types of white blood cells are identified by their microscopic appearance after histologic staining, and each has a different specialized function. The two main groups, both illustrated in [link] are the granulocytes, which include the neutrophils, eosinophils, and basophils, and the agranulocytes, which include the monocytes and lymphocytes.



(a) Granulocytes—including neutrophils, eosinophils and basophils—are characterized by a lobed nucleus and granular inclusions in the cytoplasm. Granulocytes are typically first-responders during injury or infection. (b) Agranulocytes include lymphocytes and monocytes. Lymphocytes, including B and T cells, are responsible for adaptive immune response. Monocytes differentiate into macrophages and dendritic cells, which in turn respond to infection or injury.

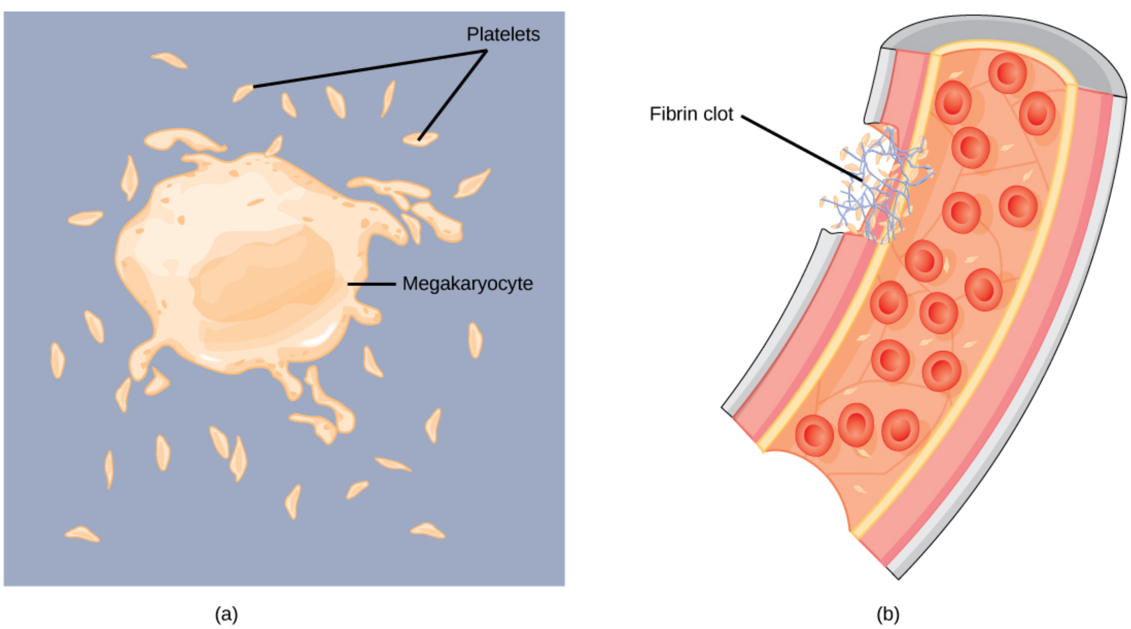
Granulocytes contain granules in their cytoplasm; the agranulocytes are so named because of the lack of granules in their cytoplasm. Some leukocytes become macrophages that either stay at the same site or move through the blood stream and gather at sites of infection or inflammation where they are attracted by chemical signals from foreign particles and damaged cells. Lymphocytes are the primary cells of the immune system and include B cells, T cells, and natural killer cells. B cells destroy bacteria and inactivate their toxins. They also produce antibodies. T cells attack viruses, fungi, some bacteria, transplanted cells, and cancer cells. T cells attack viruses by releasing toxins that kill the viruses. Natural killer cells attack a variety of infectious microbes and certain tumor cells.

One reason that HIV poses significant management challenges is because the virus directly targets T cells by gaining entry through a receptor. Once inside the cell, HIV then multiplies using the T cell’s own genetic machinery. After the HIV virus replicates, it is transmitted directly from the infected T cell to macrophages. The presence of HIV can remain

unrecognized for an extensive period of time before full disease symptoms develop.

Platelets and Coagulation Factors

Blood must clot to heal wounds and prevent excess blood loss. Small cell fragments called platelets (thrombocytes) are attracted to the wound site where they adhere by extending many projections and releasing their contents. These contents activate other platelets and also interact with other coagulation factors, which convert fibrinogen, a water-soluble protein present in blood serum into fibrin (a non-water soluble protein), causing the blood to clot. Many of the clotting factors require vitamin K to work, and vitamin K deficiency can lead to problems with blood clotting. Many platelets converge and stick together at the wound site forming a platelet plug (also called a fibrin clot), as illustrated in [link]**b**. The plug or clot lasts for a number of days and stops the loss of blood. Platelets are formed from the disintegration of larger cells called megakaryocytes, like that shown in [link]**a**. For each megakaryocyte, 2000–3000 platelets are formed with 150,000 to 400,000 platelets present in each cubic millimeter of blood. Each platelet is disc shaped and 2–4 μm in diameter. They contain many small vesicles but do not contain a nucleus.



(a) Platelets are formed from large cells called megakaryocytes. The megakaryocyte breaks up into thousands of fragments that become platelets. (b) Platelets are required for clotting of the blood. The platelets collect at a wound site in conjunction with other clotting factors, such as fibrinogen, to form a fibrin clot that prevents blood loss and allows the wound to heal.

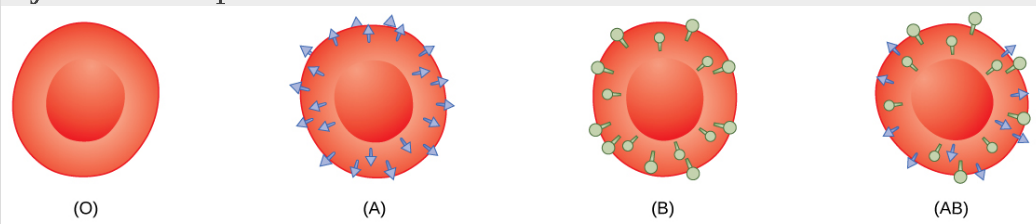
Plasma and Serum

The liquid component of blood is called plasma, and it is separated by spinning or centrifuging the blood at high rotations (3000 rpm or higher). The blood cells and platelets are separated by centrifugal forces to the bottom of a specimen tube. The upper liquid layer, the plasma, consists of 90 percent water along with various substances required for maintaining the body’s pH, osmotic load, and for protecting the body. The plasma also contains the coagulation factors and antibodies.

The plasma component of blood without the coagulation factors is called the **serum**. Serum is similar to interstitial fluid in which the correct composition of key ions acting as electrolytes is essential for normal functioning of muscles and nerves. Other components in the serum include proteins that assist with maintaining pH and osmotic balance while giving viscosity to the blood. The serum also contains antibodies, specialized proteins that are important for defense against viruses and bacteria. Lipids, including cholesterol, are also transported in the serum, along with various other substances including nutrients, hormones, metabolic waste, plus external substances, such as, drugs, viruses, and bacteria.

Human serum albumin is the most abundant protein in human blood plasma and is synthesized in the liver. Albumin, which constitutes about half of the blood serum protein, transports hormones and fatty acids, buffers pH, and maintains osmotic pressures. Immunoglobin is a protein antibody produced in the mucosal lining and plays an important role in antibody mediated immunity.

Note:
Evolution Connection
Blood Types Related to Proteins on the Surface of the Red Blood Cells
Red blood cells are coated in antigens made of glycolipids and glycoproteins. The composition of these molecules is determined by genetics, which have evolved over time. In humans, the different surface antigens are grouped into 24 different blood groups with more than 100 different antigens on each red blood cell. The two most well known blood groups are the ABO, shown in [link], and Rh systems. The surface antigens in the ABO blood group are glycolipids, called antigen A and antigen B. People with blood type A have antigen A, those with blood type B have antigen B, those with blood type AB have both antigens, and people with blood type O have neither antigen. Antibodies called agglutinougens are found in the blood plasma and react with the A or B antigens, if the two are mixed. When type A and type B blood are combined, agglutination (clumping) of the blood occurs because of antibodies in the plasma that bind with the opposing antigen; this causes clots that coagulate in the kidney causing kidney failure. Type O blood has neither A or B antigens, and therefore, type O blood can be given to all blood types. Type O negative blood is the universal donor. Type AB positive blood is the universal acceptor because it has both A and B antigen. The ABO blood groups were discovered in 1900 and 1901 by Karl Landsteiner at the University of Vienna.
The Rh blood group was first discovered in Rhesus monkeys. Most people have the Rh antigen (Rh+) and do not have anti-Rh antibodies in their blood. The few people who do not have the Rh antigen and are Rh– can develop anti-Rh antibodies if exposed to Rh+ blood. This can happen after a blood transfusion or after an Rh– woman has an Rh+ baby. The first exposure does not usually cause a reaction; however, at the second exposure, enough antibodies have built up in the blood to produce a reaction that causes agglutination and breakdown of red blood cells. An injection can prevent this reaction.



Human red blood cells may have either type A or B glycoproteins on their surface, both glycoproteins combined (AB), or neither (O). The glycoproteins serve as antigens and can elicit an immune response in a person who receives a transfusion containing unfamiliar antigens. Type O blood, which has no A or B antigens, does not elicit an immune response when injected into a person of any blood type. Thus, O is considered the universal donor. Persons with type AB blood can accept blood from any blood type, and type AB is considered the universal acceptor.

Note:
Link to Learning

Play a blood typing game on the Nobel Prize website to solidify your understanding of blood types.

Section Summary

Specific components of the blood include red blood cells, white blood cells, platelets, and the plasma, which contains coagulation factors and serum. Blood is important for regulation of the body's pH, temperature, osmotic pressure, the circulation of nutrients and removal of waste, the distribution of hormones from endocrine glands, and the elimination of excess heat; it also contains components for blood clotting. Red blood cells are specialized

cells that contain hemoglobin and circulate through the body delivering oxygen to cells. White blood cells are involved in the immune response to identify and target invading bacteria, viruses, and other foreign organisms; they also recycle waste components, such as old red blood cells. Platelets and blood clotting factors cause the change of the soluble protein fibrinogen to the insoluble protein fibrin at a wound site forming a plug. Plasma consists of 90 percent water along with various substances, such as coagulation factors and antibodies. The serum is the plasma component of the blood without the coagulation factors.

Review Questions

Exercise:

Problem: White blood cells:

a. can be classified as granulocytes or agranulocytes
b. defend the body against bacteria and viruses
c. are also called leucocytes
d. All of the above

Solution:

D

Exercise:

Problem: Platelet plug formation occurs at which point?

a. when large megakaryocytes break up into thousands of smaller fragments
b. when platelets are dispersed through the blood stream
c. when platelets are attracted to a site of blood vessel damage
d. none of the above

Solution:

C

Exercise:

Problem:

In humans, the plasma comprises what percentage of the blood?

a. 45 percent
b. 55 percent
c. 25 percent
d. 90 percent

Solution:

B

Exercise:

Problem:

The red blood cells of birds differ from mammalian red blood cells because:

a. they are white and have nuclei
b. they do not have nuclei
c. they have nuclei
d. they fight disease

Solution:

C

Free Response

Exercise:

Problem: Describe the cause of different blood type groups.

Solution:

Red blood cells are coated with proteins called antigens made of glycolipids and glycoproteins. When type A and type B blood are mixed, the blood agglutinates because of antibodies in the plasma that bind with the opposing antigen. Type O blood has no antigens. The Rh blood group has either the Rh antigen (Rh+) or no Rh antigen (Rh–).

Exercise:

Problem: List some of the functions of blood in the body.

Solution:

Blood is important for regulation of the body's pH, temperature, and osmotic pressure, the circulation of nutrients and removal of wastes, the distribution of hormones from endocrine glands, the elimination of excess heat; it also contains components for the clotting of blood to prevent blood loss. Blood also transports clotting factors and disease-fighting agents.

Exercise:

Problem: How does the lymphatic system work with blood flow?

Solution:

Lymph capillaries take fluid from the blood to the lymph nodes. The lymph nodes filter the lymph by percolation through connective tissue filled with white blood cells. The white blood cells remove infectious agents, such as bacteria and viruses, to clean the lymph before it returns to the bloodstream.

Glossary

plasma
: liquid component of blood that is left after the cells are removed

platelet
: (also, thrombocyte) small cellular fragment that collects at wounds, cross-reacts with clotting factors, and forms a plug to prevent blood loss

red blood cell
: small (7–8 μm) biconcave cell without mitochondria (and in mammals without nuclei) that is packed with hemoglobin, giving the cell its red color; transports oxygen through the body

serum
: plasma without the coagulation factors

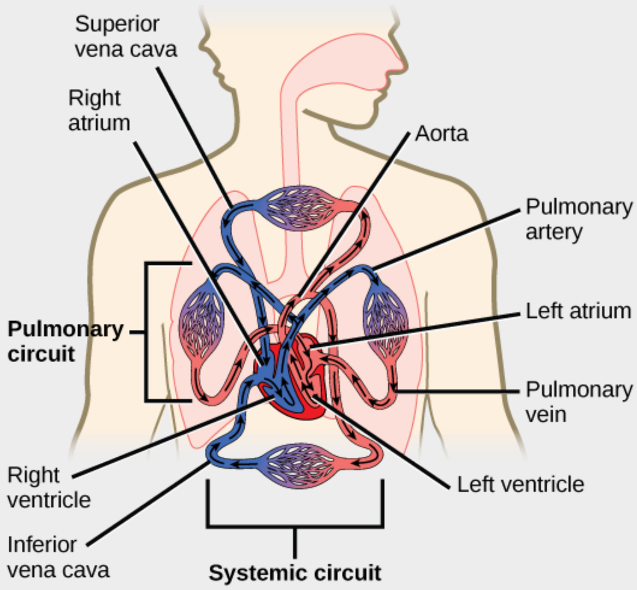
white blood cell
: large (30 μm) cell with nuclei of which there are many types with different roles including the protection of the body from viruses and bacteria, and cleaning up dead cells and other waste

Mammalian Heart and Blood Vessels
By the end of this section, you will be able to:

- Describe the structure of the heart and explain how cardiac muscle is different from other muscles
- Describe the cardiac cycle
- Explain the structure of arteries, veins, and capillaries, and how blood flows through the body

The heart is a complex muscle that pumps blood through the three divisions of the circulatory system: the coronary (vessels that serve the heart), pulmonary (heart and lungs), and systemic (systems of the body), as shown in [link]. Coronary circulation intrinsic to the heart takes blood directly from the main artery (aorta) coming from the heart. For pulmonary and systemic circulation, the heart has to pump blood to the lungs or the rest of the body, respectively. In vertebrates, the lungs are relatively close to the heart in the thoracic cavity. The shorter distance to pump means that the muscle wall on the right side of the heart is not as thick as the left side which must have enough pressure to pump blood all the way to your big toe.

Note:
Art Connection



The mammalian circulatory system is divided into three circuits: the systemic circuit, the pulmonary circuit, and the coronary circuit. Blood is pumped from veins of the systemic circuit into the right atrium of the heart, then into the right ventricle. Blood then enters the pulmonary circuit, and is oxygenated by the lungs. From the pulmonary circuit, blood re-enters the heart through the left atrium. From the left ventricle, blood re-enters the systemic circuit through the aorta and is distributed to the rest of the body. The coronary circuit, which provides blood to the heart, is not shown.

Which of the following statements about the circulatory system is false?

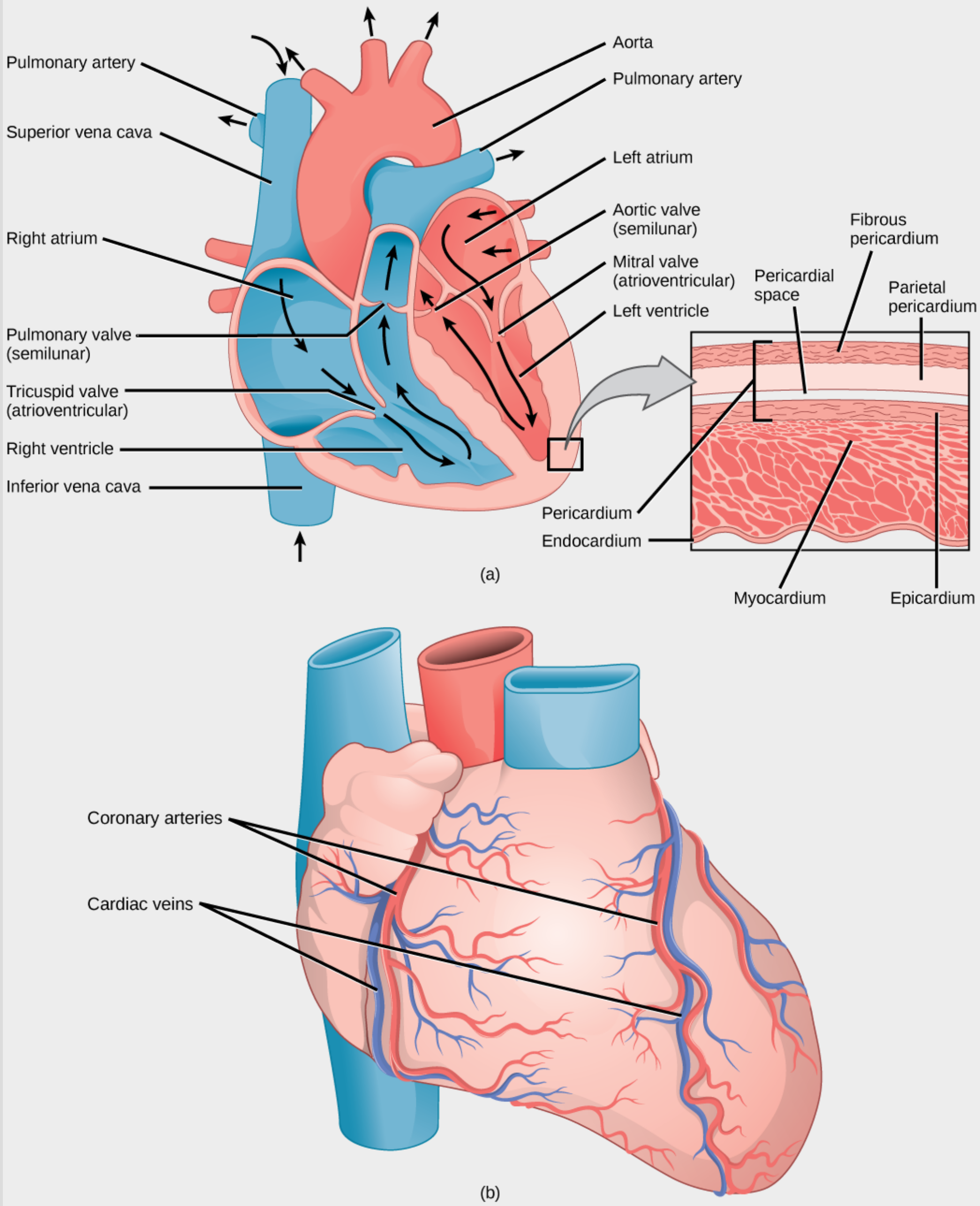
a. Blood in the pulmonary vein is deoxygenated.
b. Blood in the inferior vena cava is deoxygenated.
c. Blood in the pulmonary artery is deoxygenated.
d. Blood in the aorta is oxygenated.

Structure of the Heart

The heart muscle is asymmetrical as a result of the distance blood must travel in the pulmonary and systemic circuits. Since the right side of the heart sends blood to the pulmonary circuit it is smaller than the left side which must send blood out to the whole body in the systemic circuit, as

shown in [link]. In humans, the heart is about the size of a clenched fist; it is divided into four chambers: two atria and two ventricles. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The atria are the chambers that receive blood, and the ventricles are the chambers that pump blood. The right atrium receives deoxygenated blood from the **superior vena cava**, which drains blood from the jugular vein that comes from the brain and from the veins that come from the arms, as well as from the **inferior vena cava** which drains blood from the veins that come from the lower organs and the legs. In addition, the right atrium receives blood from the coronary sinus which drains deoxygenated blood from the heart itself. This deoxygenated blood then passes to the right ventricle through the **atrioventricular valve** or the **tricuspid valve**, a flap of connective tissue that opens in only one direction to prevent the backflow of blood. The valve separating the chambers on the left side of the heart valve is called the biscuspid or mitral valve. After it is filled, the right ventricle pumps the blood through the pulmonary arteries, by-passing the **semilunar valve** (or pulmonic valve) to the lungs for re-oxygenation. After blood passes through the pulmonary arteries, the right semilunar valves close preventing the blood from flowing backwards into the right ventricle. The left atrium then receives the oxygen-rich blood from the lungs via the pulmonary veins. This blood passes through the **bicuspid valve** or mitral valve (the atrioventricular valve on the left side of the heart) to the left ventricle where the blood is pumped out through **aorta**, the major artery of the body, taking oxygenated blood to the organs and muscles of the body. Once blood is pumped out of the left ventricle and into the aorta, the aortic semilunar valve (or aortic valve) closes preventing blood from flowing backward into the left ventricle. This pattern of pumping is referred to as double circulation and is found in all mammals.

Note:
Art Connection



(a) The heart is primarily made of a thick muscle layer, called the myocardium, surrounded by membranes. One-way valves separate the four chambers. (b) Blood vessels of the coronary system, including the coronary arteries and veins, keep the heart musculature oxygenated.

Which of the following statements about the heart is false?

a. The mitral valve separates the left ventricle from the left atrium.
b. Blood travels through the bicuspid valve to the left atrium.
c. Both the aortic and the pulmonary valves are semilunar valves.
d. The mitral valve is an atrioventricular valve.

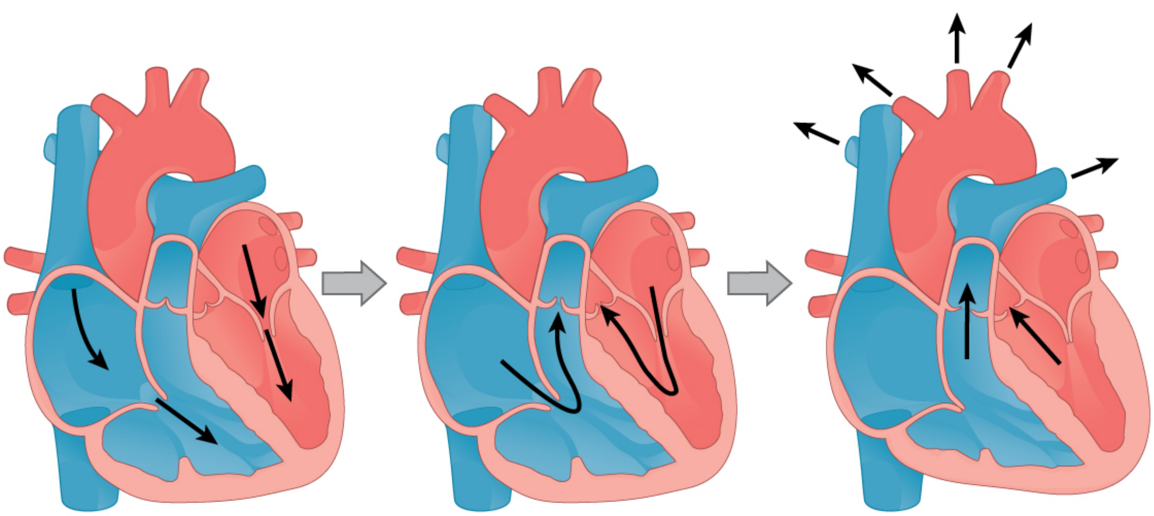
The heart is composed of three layers; the epicardium, the myocardium, and the endocardium, illustrated in [link]. The inner wall of the heart has a lining called the **endocardium**. The **myocardium** consists of the heart muscle cells that make up the middle layer and the bulk of the heart wall. The outer layer of cells is called the **epicardium**, of which the second layer is a membranous layered structure called the **pericardium** that surrounds and protects the heart; it allows enough room for vigorous pumping but also keeps the heart in place to reduce friction between the heart and other structures.

The heart has its own blood vessels that supply the heart muscle with blood. The **coronary arteries** branch from the aorta and surround the outer surface of the heart like a crown. They diverge into capillaries where the heart muscle is supplied with oxygen before converging again into the **coronary veins** to take the deoxygenated blood back to the right atrium where the blood will be re-oxygenated through the pulmonary circuit. The heart muscle will die without a steady supply of blood. **Atherosclerosis** is the blockage of an artery by the buildup of fatty plaques. Because of the size (narrow) of the coronary arteries and their function in serving the heart itself, atherosclerosis can be deadly in these arteries. The slowdown of blood flow and subsequent oxygen deprivation that results from atherosclerosis causes severe pain, known as **angina**, and complete blockage of the arteries will cause **myocardial infarction**: the death of cardiac muscle tissue, commonly known as a heart attack.

The Cardiac Cycle

The main purpose of the heart is to pump blood through the body; it does so in a repeating sequence called the cardiac cycle. The **cardiac cycle** is the

coordination of the filling and emptying of the heart of blood by electrical signals that cause the heart muscles to contract and relax. The human heart beats over 100,000 times per day. In each cardiac cycle, the heart contracts (**systole**), pushing out the blood and pumping it through the body; this is followed by a relaxation phase (**diastole**), where the heart fills with blood, as illustrated in [link]. The atria contract at the same time, forcing blood through the atrioventricular valves into the ventricles. Closing of the atrioventricular valves produces a monosyllabic “lup” sound. Following a brief delay, the ventricles contract at the same time forcing blood through the semilunar valves into the aorta and the artery transporting blood to the lungs (via the pulmonary artery). Closing of the semilunar valves produces a monosyllabic “dup” sound.



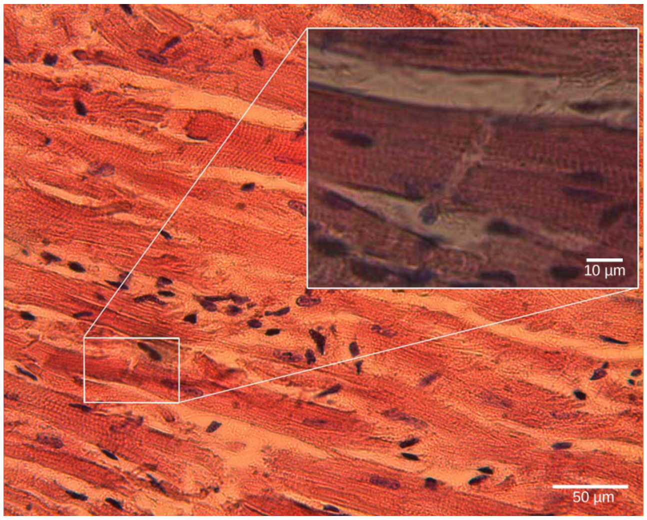
(a) Cardiac diastole: all chambers are relaxed, and blood flows into the heart.

(b) Atrial systole, ventricular diastole: atria contract, pushing blood into the ventricles.

(c) Atrial diastole, ventricular systole: after the atria relax, the ventricles contract, pushing blood out of the heart.

During (a) cardiac diastole, the heart muscle is relaxed and blood flows into the heart. During (b) atrial systole, the atria contract, pushing blood into the ventricles. During (c) atrial diastole, the ventricles contract, forcing blood out of the heart.

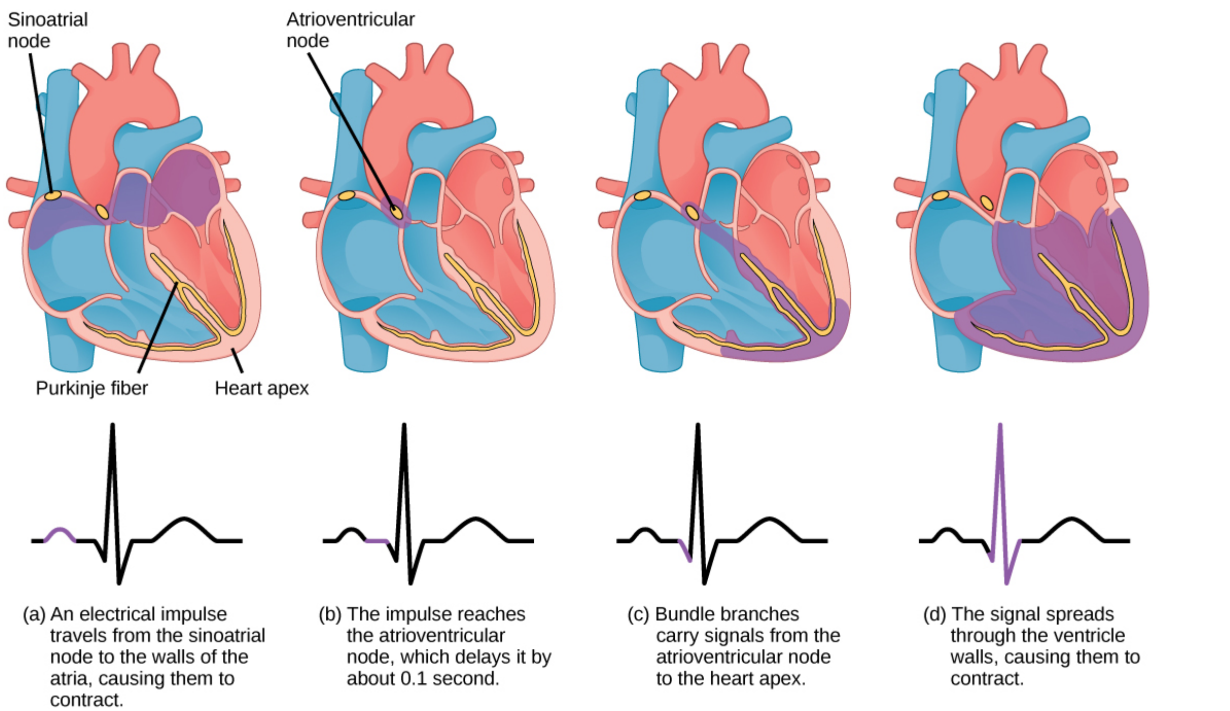
The pumping of the heart is a function of the cardiac muscle cells, or cardiomyocytes, that make up the heart muscle. **Cardiomyocytes**, shown in [link], are distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle; they are connected by intercalated disks exclusive to cardiac muscle. They are self-stimulated for a period of time and isolated cardiomyocytes will beat if given the correct balance of nutrients and electrolytes.



Cardiomyocytes are striated muscle cells found in cardiac tissue. (credit: modification of work by Dr. S. Girod, Anton Becker; scale-bar data from Matt Russell)

The autonomous beating of cardiac muscle cells is regulated by the heart's internal pacemaker that uses electrical signals to time the beating of the heart. The electrical signals and mechanical actions, illustrated in [link], are intimately intertwined. The internal pacemaker starts at the **sinoatrial (SA) node**, which is located near the wall of the right atrium. Electrical charges spontaneously pulse from the SA node causing the two atria to contract in

unison. The pulse reaches a second node, called the atrioventricular (AV) node, between the right atrium and right ventricle where it pauses for approximately 0.1 second before spreading to the walls of the ventricles. From the AV node, the electrical impulse enters the bundle of His, then to the left and right bundle branches extending through the interventricular septum. Finally, the Purkinje fibers conduct the impulse from the apex of the heart up the ventricular myocardium, and then the ventricles contract. This pause allows the atria to empty completely into the ventricles before the ventricles pump out the blood. The electrical impulses in the heart produce electrical currents that flow through the body and can be measured on the skin using electrodes. This information can be observed as an **electrocardiogram (ECG)**—a recording of the electrical impulses of the cardiac muscle.



The beating of the heart is regulated by an electrical impulse that causes the characteristic reading of an ECG. The signal is initiated at the sinoatrial valve. The signal then (a) spreads to the atria, causing them to contract. The signal is (b) delayed at the atrioventricular node before it is passed on to the (c) heart apex.

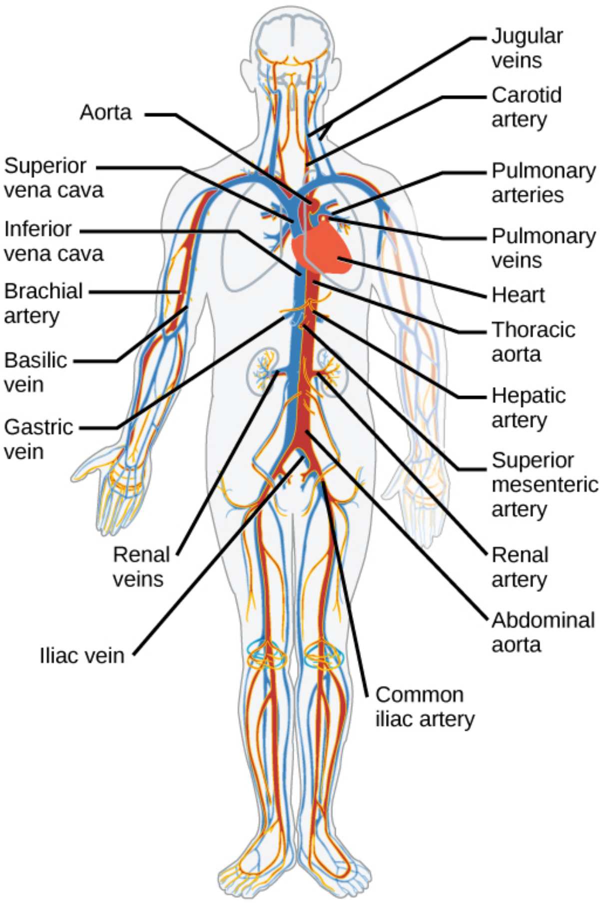
The delay allows the atria to relax before the (d) ventricles contract. The final part of the ECG cycle prepares the heart for the next beat.

Note:
Link to Learning

Visit this site and select the dropdown "Your Heart's Electrical System" to see the heart's "pacemaker" in action.

Arteries, Veins, and Capillaries

The blood from the heart is carried through the body by a complex network of blood vessels ([link]). **Arteries** take blood away from the heart. The main artery is the aorta that branches into major arteries that take blood to different limbs and organs. These major arteries include the carotid artery that takes blood to the brain, the brachial arteries that take blood to the arms, and the thoracic artery that takes blood to the thorax and then into the hepatic, renal, and gastric arteries for the liver, kidney, and stomach, respectively. The iliac artery takes blood to the lower limbs. The major arteries diverge into minor arteries, and then smaller vessels called **arterioles**, to reach more deeply into the muscles and organs of the body.



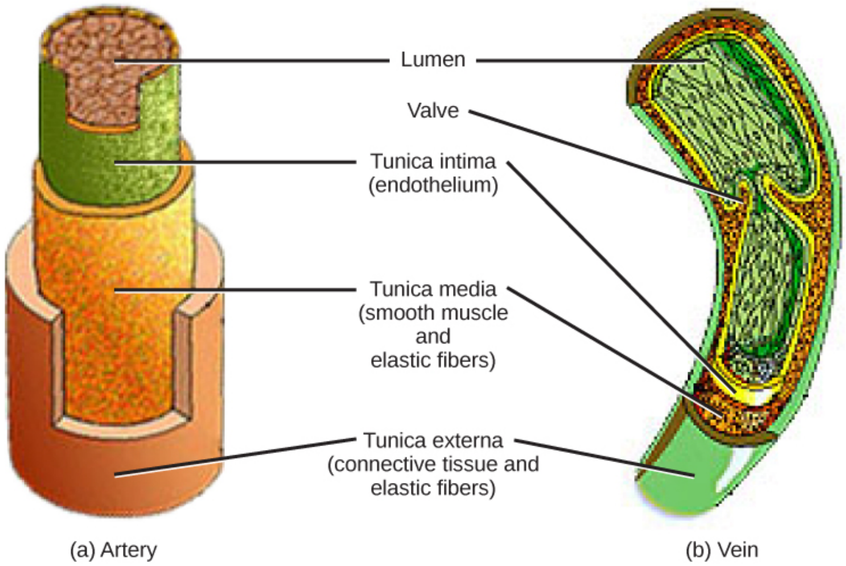
The major human arteries and veins are shown. (credit: modification of work by Mariana Ruiz Villareal)

Arterioles diverge into capillary beds. **Capillary beds** contain a large number (10 to 100) of **capillaries** that branch among the cells and tissues of the body. Capillaries are narrow-diameter tubes that can fit red blood cells through in single file and are the sites for the exchange of nutrients, waste, and oxygen with tissues at the cellular level. Fluid also crosses into the interstitial space from the capillaries. The capillaries converge again into **venules** that connect to minor veins that finally connect to major veins that take blood high in carbon dioxide back to the heart. **Veins** are blood vessels that bring blood back to the heart. The major veins drain blood from the

same organs and limbs that the major arteries supply. Fluid is also brought back to the heart via the lymphatic system.

The structure of the different types of blood vessels reflects their function or layers. There are three distinct layers, or tunics, that form the walls of blood vessels ([link]). The first tunic is a smooth, inner lining of endothelial cells that are in contact with the red blood cells. The endothelial tunic is continuous with the endocardium of the heart. In capillaries, this single layer of cells is the location of diffusion of oxygen and carbon dioxide between the endothelial cells and red blood cells, as well as the exchange site via endocytosis and exocytosis. The movement of materials at the site of capillaries is regulated by **vasoconstriction**, narrowing of the blood vessels, and **vasodilation**, widening of the blood vessels; this is important in the overall regulation of blood pressure.

Veins and arteries both have two further tunics that surround the endothelium: the middle tunic is composed of smooth muscle and the outermost layer is connective tissue (collagen and elastic fibers). The elastic connective tissue stretches and supports the blood vessels, and the smooth muscle layer helps regulate blood flow by altering vascular resistance through vasoconstriction and vasodilation. The arteries have thicker smooth muscle and connective tissue than the veins to accommodate the higher pressure and speed of freshly pumped blood. The veins are thinner walled as the pressure and rate of flow are much lower. In addition, veins are structurally different than arteries in that veins have valves to prevent the backflow of blood. Because veins have to work against gravity to get blood back to the heart, contraction of skeletal muscle assists with the flow of blood back to the heart.



Arteries and veins consist of three layers: an outer tunica externa, a middle tunica media, and an inner tunica intima. Capillaries consist of a single layer of epithelial cells, the tunica intima. (credit: modification of work by NCI, NIH)

Section Summary

The heart muscle pumps blood through three divisions of the circulatory system: coronary, pulmonary, and systemic. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The pumping of the heart is a function of cardiomyocytes, distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle. The internal pacemaker starts at the sinoatrial node, which is located near the wall of the right atrium. Electrical charges pulse from the SA node causing the two atria to contract in unison; then the pulse reaches the atrioventricular node between the right atrium and right ventricle. A pause in the electric signal allows the atria to empty completely into the ventricles before the ventricles pump out the blood. The blood from the heart is carried through the body by a complex network of

blood vessels; arteries take blood away from the heart, and veins bring blood back to the heart.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the circulatory system is false?

a. Blood in the pulmonary vein is deoxygenated.
b. Blood in the inferior vena cava is deoxygenated.
c. Blood in the pulmonary artery is deoxygenated.
d. Blood in the aorta is oxygenated.

Solution:

[link] C

Exercise:

Problem:

[link] Which of the following statements about the heart is false?

a. The mitral valve separates the left ventricle from the left atrium.
b. Blood travels through the bicuspid valve to the left atrium.
c. Both the aortic and the pulmonary valves are semilunar valves.
d. The mitral valve is an atrioventricular valve.

Solution:

[link] B

Review Questions

Exercise:

Problem: The heart’s internal pacemaker beats by:

a. an internal implant that sends an electrical impulse through the heart
b. the excitation of cardiac muscle cells at the sinoatrial node followed by the atrioventricular node
c. the excitation of cardiac muscle cells at the atrioventricular node followed by the sinoatrial node
d. the action of the sinus

Solution:

B

Exercise:

Problem:

During the systolic phase of the cardiac cycle, the heart is _________.

a. contracting
b. relaxing
c. contracting and relaxing
d. filling with blood

Solution:

A

Exercise:

Problem: Cardiomyocytes are similar to skeletal muscle because:

a. they beat involuntarily
b. they are used for weight lifting

c. they pulse rhythmically
d. they are striated

Solution:

D

Exercise:

Problem: How do arteries differ from veins?

a. Arteries have thicker smooth muscle layers to accommodate the changes in pressure from the heart.
b. Arteries carry blood.
c. Arteries have thinner smooth muscle layers and valves and move blood by the action of skeletal muscle.
d. Arteries are thin walled and are used for gas exchange.

Solution:

A

Free Response

Exercise:

Problem: Describe the cardiac cycle.

Solution:

The heart receives an electrical signal from the sinoatrial node triggering the cardiac muscle cells in the atria to contract. The signal pauses at the atrioventricular node before spreading to the walls of the ventricles so the blood is pumped through the body. This is the systolic phase. The heart then relaxes in the diastole and fills again with blood.

Exercise:

Problem: What happens in capillaries?

Solution:

The capillaries basically exchange materials with their surroundings. Their walls are very thin and are made of one or two layers of cells, where gases, nutrients, and waste are diffused. They are distributed as beds, complex networks that link arteries as well as veins.

Glossary

angina
: pain caused by partial blockage of the coronary arteries by the buildup of plaque and lack of oxygen to the heart muscle

aorta
: major artery of the body that takes blood away from the heart

arteriole
: small vessel that connects an artery to a capillary bed

artery
: blood vessel that takes blood away from the heart

atherosclerosis
: buildup of fatty plaques in the coronary arteries in the heart

atrioventricular valve
: one-way membranous flap of connective tissue between the atrium and the ventricle in the right side of the heart; also known as tricuspid valve

bicuspid valve
: (also, mitral valve; left atrioventricular valve) one-way membranous flap between the atrium and the ventricle in the left side of the heart

capillary

smallest blood vessel that allows the passage of individual blood cells and the site of diffusion of oxygen and nutrient exchange

capillary bed

large number of capillaries that converge to take blood to a particular organ or tissue

cardiac cycle

filling and emptying the heart of blood by electrical signals that cause the heart muscles to contract and relax

cardiomyocyte

specialized heart muscle cell that is striated but contracts involuntarily like smooth muscle

coronary artery

vessel that supplies the heart tissue with blood

coronary vein

vessel that takes blood away from the heart tissue back to the chambers in the heart

diastole

relaxation phase of the cardiac cycle when the heart is relaxed and the ventricles are filling with blood

electrocardiogram (ECG)

recording of the electrical impulses of the cardiac muscle

endocardium

innermost layer of tissue in the heart

epicardium

outermost tissue layer of the heart

inferior vena cava

drains blood from the veins that come from the lower organs and the legs

myocardial infarction
(also, heart attack) complete blockage of the coronary arteries and death of the cardiac muscle tissue

myocardium
heart muscle cells that make up the middle layer and the bulk of the heart wall

pericardium
membrane layer protecting the heart; also part of the epicardium

semilunar valve
membranous flap of connective tissue between the aorta and a ventricle of the heart (the aortic or pulmonary semilunar valves)

sinoatrial (SA) node
the heart’s internal pacemaker; located near the wall of the right atrium

superior vena cava
drains blood from the jugular vein that comes from the brain and from the veins that come from the arms

systole
contraction phase of cardiac cycle when the ventricles are pumping blood into the arteries

tricuspid valve
one-way membranous flap of connective tissue between the atrium and the ventricle in the right side of the heart; also known as atrioventricular valve

vasoconstriction
narrowing of a blood vessel

vasodilation

widening of a blood vessel

vein
: blood vessel that brings blood back to the heart

vena cava
: major vein of the body returning blood from the upper and lower parts of the body; see the superior vena cava and inferior vena cava

venule
: blood vessel that connects a capillary bed to a vein

Blood Flow and Blood Pressure Regulation
By the end of this section, you will be able to:

- Describe the system of blood flow through the body
- Describe how blood pressure is regulated

Blood pressure (BP) is the pressure exerted by blood on the walls of a blood vessel that helps to push blood through the body. Systolic blood pressure measures the amount of pressure that blood exerts on vessels while the heart is beating. The optimal systolic blood pressure is 120 mmHg. Diastolic blood pressure measures the pressure in the vessels between heartbeats. The optimal diastolic blood pressure is 80 mmHg. Many factors can affect blood pressure, such as hormones, stress, exercise, eating, sitting, and standing. Blood flow through the body is regulated by the size of blood vessels, by the action of smooth muscle, by one-way valves, and by the fluid pressure of the blood itself.

How Blood Flows Through the Body

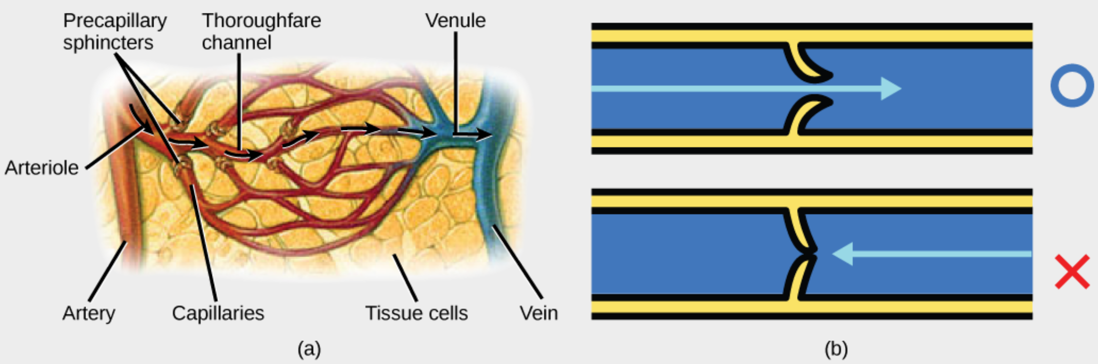
Blood is pushed through the body by the action of the pumping heart. With each rhythmic pump, blood is pushed under high pressure and velocity away from the heart, initially along the main artery, the aorta. In the aorta, the blood travels at 30 cm/sec. As blood moves into the arteries, arterioles, and ultimately to the capillary beds, the rate of movement slows dramatically to about 0.026 cm/sec, one-thousand times slower than the rate of movement in the aorta. While the diameter of each individual arteriole and capillary is far narrower than the diameter of the aorta, and according to the law of continuity, fluid should travel faster through a narrower diameter tube, the rate is actually slower due to the overall diameter of all the combined capillaries being far greater than the diameter of the individual aorta.

The slow rate of travel through the capillary beds, which reach almost every cell in the body, assists with gas and nutrient exchange and also promotes the diffusion of fluid into the interstitial space. After the blood has passed through the capillary beds to the venules, veins, and finally to the main venae cavae, the rate of flow increases again but is still much slower than

the initial rate in the aorta. Blood primarily moves in the veins by the rhythmic movement of smooth muscle in the vessel wall and by the action of the skeletal muscle as the body moves. Because most veins must move blood against the pull of gravity, blood is prevented from flowing backward in the veins by one-way valves. Because skeletal muscle contraction aids in venous blood flow, it is important to get up and move frequently after long periods of sitting so that blood will not pool in the extremities.

Blood flow through the capillary beds is regulated depending on the body's needs and is directed by nerve and hormone signals. For example, after a large meal, most of the blood is diverted to the stomach by vasodilation of vessels of the digestive system and vasoconstriction of other vessels. During exercise, blood is diverted to the skeletal muscles through vasodilation while blood to the digestive system would be lessened through vasoconstriction. The blood entering some capillary beds is controlled by small muscles, called precapillary sphincters, illustrated in [link]. If the sphincters are open, the blood will flow into the associated branches of the capillary blood. If all of the sphincters are closed, then the blood will flow directly from the arteriole to the venule through the thoroughfare channel (see [link]). These muscles allow the body to precisely control when capillary beds receive blood flow. At any given moment only about 5-10% of our capillary beds actually have blood flowing through them.

Note:
Art Connection



(a) Precapillary sphincters are rings of smooth muscle that regulate the flow of blood through capillaries; they help control

the location of blood flow to where it is needed. (b) Valves in the veins prevent blood from moving backward. (credit a: modification of work by NCI)

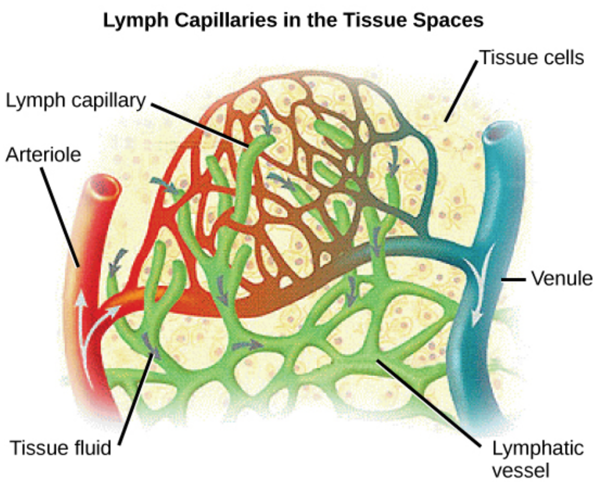
Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?

Note:
Link to Learning

See the circulatory system's blood flow.
https://www.openstaxcollege.org/l/circulation

Proteins and other large solutes cannot leave the capillaries. The loss of the watery plasma creates a hyperosmotic solution within the capillaries, especially near the venules. This causes about 85% of the plasma that leaves the capillaries to eventually diffuses back into the capillaries near the venules. The remaining 15% of blood plasma drains out from the interstitial fluid into nearby lymphatic vessels ([link]). The fluid in the lymph is similar in composition to the interstitial fluid. The lymph fluid passes through lymph nodes before it returns to the heart via the vena cava. **Lymph nodes** are specialized organs that filter the lymph by percolation through a maze of connective tissue filled with white blood cells. The white blood cells remove infectious agents, such as bacteria and viruses, to clean the lymph before it returns to the bloodstream. After it is cleaned, the lymph

returns to the heart by the action of smooth muscle pumping, skeletal muscle action, and one-way valves joining the returning blood near the junction of the venae cavae entering the right atrium of the heart.



Fluid from the capillaries moves into the interstitial space and lymph capillaries by diffusion down a pressure gradient and also by osmosis. Out of 7,200 liters of fluid pumped by the average heart in a day, over 1,500 liters is filtered. (credit: modification of work by NCI, NIH)

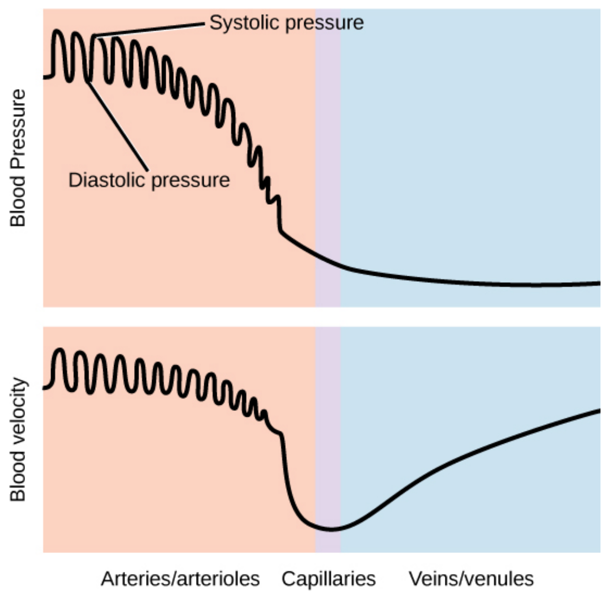
Note:
Evolution Connection
Vertebrate Diversity in Blood Circulation
Blood circulation has evolved differently in vertebrates and may show variation in different animals for the required amount of pressure, organ

and vessel location, and organ size. Animals with longs necks and those that live in cold environments have distinct blood pressure adaptations. Long necked animals, such as giraffes, need to pump blood upward from the heart against gravity. The blood pressure required from the pumping of the left ventricle would be equivalent to 250 mm Hg (mm Hg = millimeters of mercury, a unit of pressure) to reach the height of a giraffe's head, which is 2.5 meters higher than the heart. However, if checks and balances were not in place, this blood pressure would damage the giraffe's brain, particularly if it was bending down to drink. These checks and balances include valves and feedback mechanisms that reduce the rate of cardiac output. Long-necked dinosaurs such as the sauropods had to pump blood even higher, up to ten meters above the heart. This would have required a blood pressure of more than 600 mm Hg, which could only have been achieved by an enormous heart. Evidence for such an enormous heart does not exist and mechanisms to reduce the blood pressure required include the slowing of metabolism as these animals grew larger. It is likely that they did not routinely feed on tree tops but grazed on the ground.
Living in cold water, whales need to maintain the temperature in their blood. This is achieved by the veins and arteries being close together so that heat exchange can occur. This mechanism is called a countercurrent heat exchanger. The blood vessels and the whole body are also protected by thick layers of blubber to prevent heat loss. In land animals that live in cold environments, thick fur and hibernation are used to retain heat and slow metabolism.

Blood Pressure

The pressure of the blood flow in the body is produced by the hydrostatic pressure of the fluid (blood) against the walls of the blood vessels. Fluid will move from areas of high to low hydrostatic pressures. In the arteries, the hydrostatic pressure near the heart is very high and blood flows to the arterioles where the rate of flow is slowed by the narrow openings of the arterioles. During systole, when new blood is entering the arteries, the artery walls stretch to accommodate the increase of pressure of the extra blood; during diastole, the walls return to normal because of their elastic

properties. The blood pressure of the systole phase and the diastole phase, graphed in [link], gives the two pressure readings for blood pressure. For example, 120/80 indicates a reading of 120 mm Hg during the systole and 80 mm Hg during diastole. Throughout the cardiac cycle, the blood continues to empty into the arterioles at a relatively even rate. This resistance to blood flow is called **peripheral resistance**.



Blood pressure is related to the blood velocity in the arteries and arterioles. In the capillaries and veins, the blood pressure continues to decease but velocity increases.

Blood Pressure Regulation

Cardiac output is the volume of blood pumped by the heart in one minute. It is calculated by multiplying the number of heart contractions that occur per minute (heart rate) times the **stroke volume** (the volume of blood pumped

into the aorta per contraction of the left ventricle). Therefore, cardiac output can be increased by increasing heart rate, as when exercising. However, cardiac output can also be increased by increasing stroke volume, such as if the heart contracts with greater strength. Stroke volume can also be increased by speeding blood circulation through the body so that more blood enters the heart between contractions. During heavy exertion, the blood vessels relax and increase in diameter, offsetting the increased heart rate and ensuring adequate oxygenated blood gets to the muscles. Stress triggers a decrease in the diameter of the blood vessels, consequently increasing blood pressure. These changes can also be caused by nerve signals or hormones, and even standing up or lying down can have a great effect on blood pressure.

Section Summary

Blood primarily moves through the body by the rhythmic movement of smooth muscle in the vessel wall and by the action of the skeletal muscle as the body moves. Blood is prevented from flowing backward in the veins by one-way valves. Blood flow through the capillary beds is controlled by precapillary sphincters to increase and decrease flow depending on the body’s needs and is directed by nerve and hormone signals. Lymph vessels take fluid that has leaked out of the blood to the lymph nodes where it is cleaned before returning to the heart. During systole, blood enters the arteries, and the artery walls stretch to accommodate the extra blood. During diastole, the artery walls return to normal. The blood pressure of the systole phase and the diastole phase gives the two pressure readings for blood pressure.

Art Connections

Exercise:

Problem:

[link] Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?

Solution:

[link] Blood in the legs is farthest away from the heart and has to flow up to reach it.

Review Questions

Exercise:

Problem: High blood pressure would be a result of ________.

a. a high cardiac output and high peripheral resistance
b. a high cardiac output and low peripheral resistance
c. a low cardiac output and high peripheral resistance
d. a low cardiac output and low peripheral resistance

Solution:

A

Free Response

Exercise:

Problem: How does blood pressure change during heavy exercise?

Solution:

The heart rate increases, which increases the hydrostatic pressure against the artery walls. At the same time, the arterioles dilate in response to the increased exercise, which reduces peripheral resistance.

Glossary

blood pressure (BP)
: pressure of blood in the arteries that helps to push blood through the body

cardiac output
: the volume of blood pumped by the heart in one minute as a product of heart rate multiplied by stroke volume

lymph node
: specialized organ that contains a large number of macrophages that clean the lymph before the fluid is returned to the heart

peripheral resistance
: resistance of the artery and blood vessel walls to the pressure placed on them by the force of the heart pumping

precapillary sphincter
: small muscle that controls blood circulation in the capillary beds

stroke volume>
: - the volume of blood pumped into the aorta per contraction of the left ventricle

Systems of Gas Exchange

By the end of this section, you will be able to:

- Describe the passage of air from the outside environment to the lungs
- Explain how the lungs are protected from particulate matter

The primary function of the respiratory system is to deliver oxygen to the cells of the body’s tissues and remove carbon dioxide, a cell waste product. The main structures of the human respiratory system are the nasal cavity, the trachea, and lungs.

All aerobic organisms require oxygen to carry out their metabolic functions. Along the evolutionary tree, different organisms have devised different means of obtaining oxygen from the surrounding atmosphere. The environment in which the animal lives greatly determines how an animal respires. The complexity of the respiratory system is correlated with the size of the organism. As animal size increases, diffusion distances increase and the ratio of surface area to volume drops. In unicellular organisms, diffusion across the cell membrane is sufficient for supplying oxygen to the cell ([link]). Diffusion is a slow, passive transport process. In order for diffusion to be a feasible means of providing oxygen to the cell, the rate of oxygen uptake must match the rate of diffusion across the membrane. In other words, if the cell were very large or thick, diffusion would not be able to provide oxygen quickly enough to the inside of the cell. Therefore, dependence on diffusion as a means of obtaining oxygen and removing carbon dioxide remains feasible only for small organisms or those with highly-flattened bodies, such as many flatworms (Platyhelminthes). Larger organisms had to evolve specialized respiratory tissues, such as gills, lungs, and respiratory passages accompanied by complex circulatory systems, to transport oxygen throughout their entire body.

The cell of the unicellular algae *Ventricaria ventricosa* is one of the largest known, reaching one to five centimeters in diameter. Like all single-celled organisms, *V. ventricosa* exchanges gases across the cell membrane.

Direct Diffusion

For small multicellular organisms, diffusion across the outer membrane is sufficient to meet their oxygen needs. Gas exchange by direct diffusion across surface membranes is efficient for organisms less than 1 mm in diameter. In simple organisms, such as cnidarians and flatworms, every cell in the body is close to the external environment. Their cells are kept moist and gases diffuse quickly via direct diffusion. Flatworms are small, literally flat worms, which ‘breathe’ through diffusion across the outer membrane ([link]). The flat shape of these organisms increases the surface area for diffusion, ensuring that each cell within the body is close to the outer membrane surface and has access to oxygen. If the flatworm had a

cylindrical body, then the cells in the center would not be able to get oxygen.

This flatworm's process of respiration works by diffusion across the outer membrane. (credit: Stephen Childs)

Skin and Gills

Earthworms and amphibians use their skin (integument) as a respiratory organ. A dense network of capillaries lies just below the skin and facilitates gas exchange between the external environment and the circulatory system. The respiratory surface must be kept moist in order for the gases to dissolve and diffuse across cell membranes.

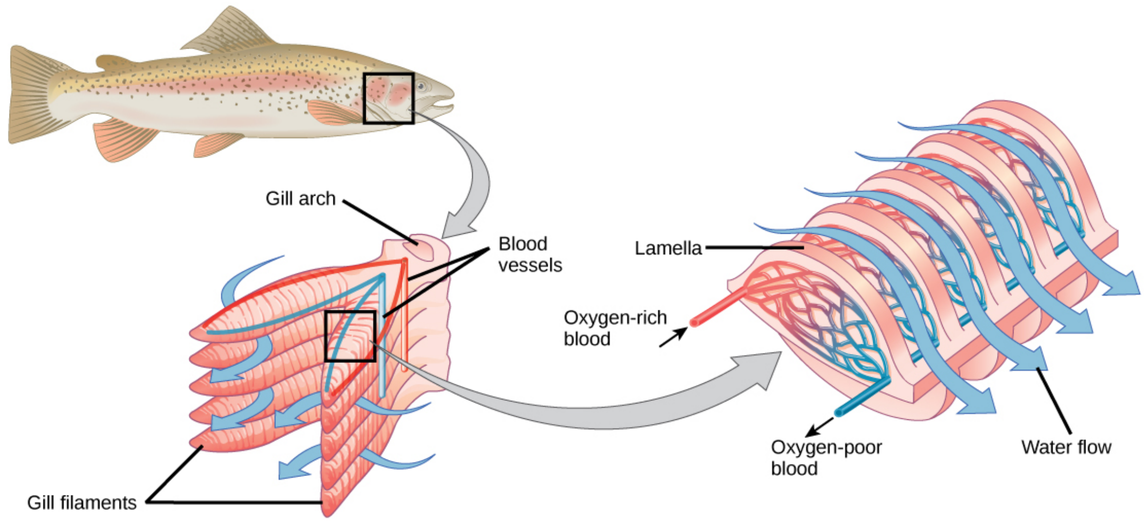
Organisms that live in water need to obtain oxygen from the water. Oxygen dissolves in water but at a lower concentration than in the atmosphere. The atmosphere has roughly 21 percent oxygen. In water, the oxygen concentration is much smaller than that. Fish and many other aquatic organisms have evolved gills to take up the dissolved oxygen from water ([link]). Gills are thin tissue filaments that are highly branched and folded. When water passes over the gills, the dissolved oxygen in water rapidly

diffuses across the gills into the bloodstream. The circulatory system can then carry the oxygenated blood to the other parts of the body. In animals that contain coelomic fluid instead of blood, oxygen diffuses across the gill surfaces into the coelomic fluid. Gills are found in mollusks, annelids, and crustaceans.

This common carp, like many other aquatic organisms, has gills that allow it to obtain oxygen from water. (credit: "Guitardude012"/Wikimedia Commons)

The folded surfaces of the gills provide a large surface area to ensure that the fish gets sufficient oxygen. Diffusion is a process in which material travels from regions of high concentration to low concentration until equilibrium is reached. In this case, blood with a low concentration of oxygen molecules circulates through the gills. The concentration of oxygen molecules in water is higher than the concentration of oxygen molecules in gills. As a result, oxygen molecules diffuse from water (high concentration) to blood (low concentration), as shown in [link]. Similarly, carbon dioxide

molecules in the blood diffuse from the blood (high concentration) to water (low concentration).



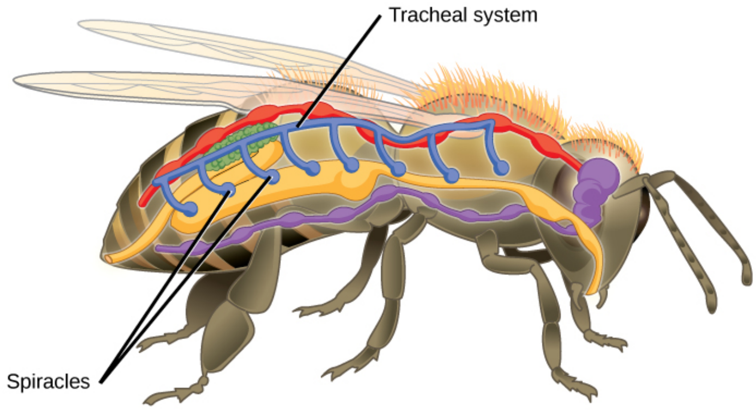
As water flows over the gills, oxygen is transferred to blood via the veins. (credit "fish": modification of work by Duane Raver, NOAA)

Tracheal Systems

Insect respiration is independent of its circulatory system; therefore, the blood does not play a direct role in oxygen transport. Insects have a highly specialized type of respiratory system called the tracheal system, which consists of a network of small tubes that carries oxygen to the entire body. The tracheal system is the most direct and efficient respiratory system in active animals. The tubes in the tracheal system are made of a polymeric material called chitin.

Insect bodies have openings, called spiracles, along the thorax and abdomen. These openings connect to the tubular network, allowing oxygen to pass into the body ([link]) and regulating the diffusion of CO_2 and water

vapor. Air enters and leaves the tracheal system through the spiracles. Some insects can ventilate the tracheal system with body movements.



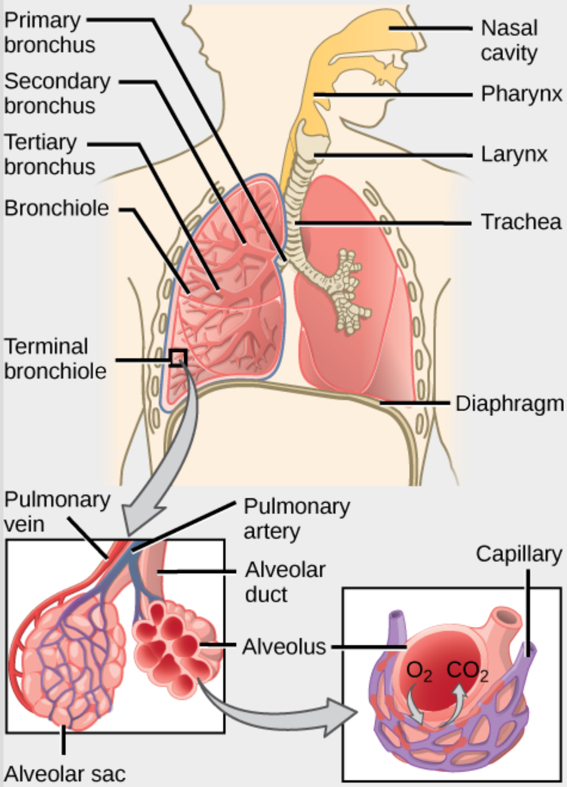
Insects perform respiration via a tracheal system.

Mammalian Systems

In mammals, pulmonary ventilation occurs via inhalation (breathing). During inhalation, air enters the body through the **nasal cavity** located just inside the nose ([link]). As air passes through the nasal cavity, the air is warmed to body temperature and humidified. The respiratory tract is coated with mucus to seal the tissues from direct contact with air. Mucus is high in water. As air crosses these surfaces of the mucous membranes, it picks up water. These processes help equilibrate the air to the body conditions, reducing any damage that cold, dry air can cause. Particulate matter that is floating in the air is removed in the nasal passages via mucus and cilia. The processes of warming, humidifying, and removing particles are important protective mechanisms that prevent damage to the trachea and lungs. Thus, inhalation serves several purposes in addition to bringing oxygen into the respiratory system.

Note:

Art Connection

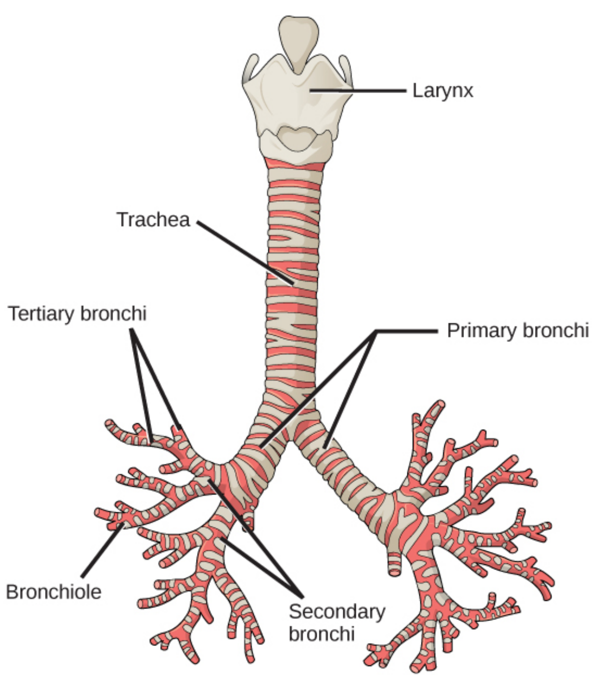


Air enters the respiratory system through the nasal cavity and pharynx, and then passes through the trachea and into the bronchi, which bring air into the lungs. (credit: modification of work by NCI)

Which of the following statements about the mammalian respiratory system is false?

a. When we breathe in, air travels from the pharynx to the trachea.
b. The bronchioles branch into bronchi.
c. Alveolar ducts connect to alveolar sacs.
d. Gas exchange between the lung and blood takes place in the alveolus.

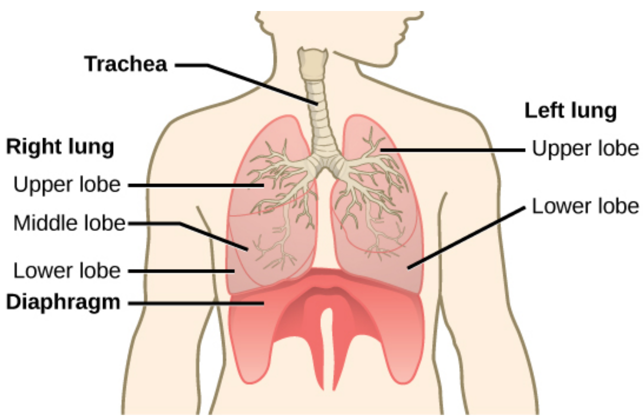
From the nasal cavity, air passes through the **pharynx** (throat) and the **larynx** (voice box), as it makes its way to the **trachea** ([link]). The main function of the trachea is to funnel the inhaled air to the lungs and the exhaled air back out of the body. The human trachea is a cylinder about 10 to 12 cm long and 2 cm in diameter that sits in front of the esophagus and extends from the larynx into the chest cavity where it divides into the two primary bronchi at the midthorax. It is made of incomplete rings of hyaline cartilage and smooth muscle ([link]). The trachea is lined with mucus-producing goblet cells and ciliated epithelia. The cilia propel foreign particles trapped in the mucus toward the pharynx. The cartilage provides strength and support to the trachea to keep the passage open. The smooth muscle can contract, decreasing the trachea's diameter, which causes expired air to rush upwards from the lungs at a great force. The forced exhalation helps expel mucus when we cough. Smooth muscle can contract or relax, depending on stimuli from the external environment or the body's nervous system.



The trachea and bronchi are made of incomplete rings of cartilage. (credit: modification of work by Gray's Anatomy)

Lungs: Bronchi and Alveoli

The end of the trachea bifurcates (divides) to the right and left lungs. The lungs are not identical. The right lung is larger and contains three lobes, whereas the smaller left lung contains two lobes ([link]). The muscular **diaphragm**, which facilitates breathing, is inferior to (below) the lungs and marks the end of the thoracic cavity.

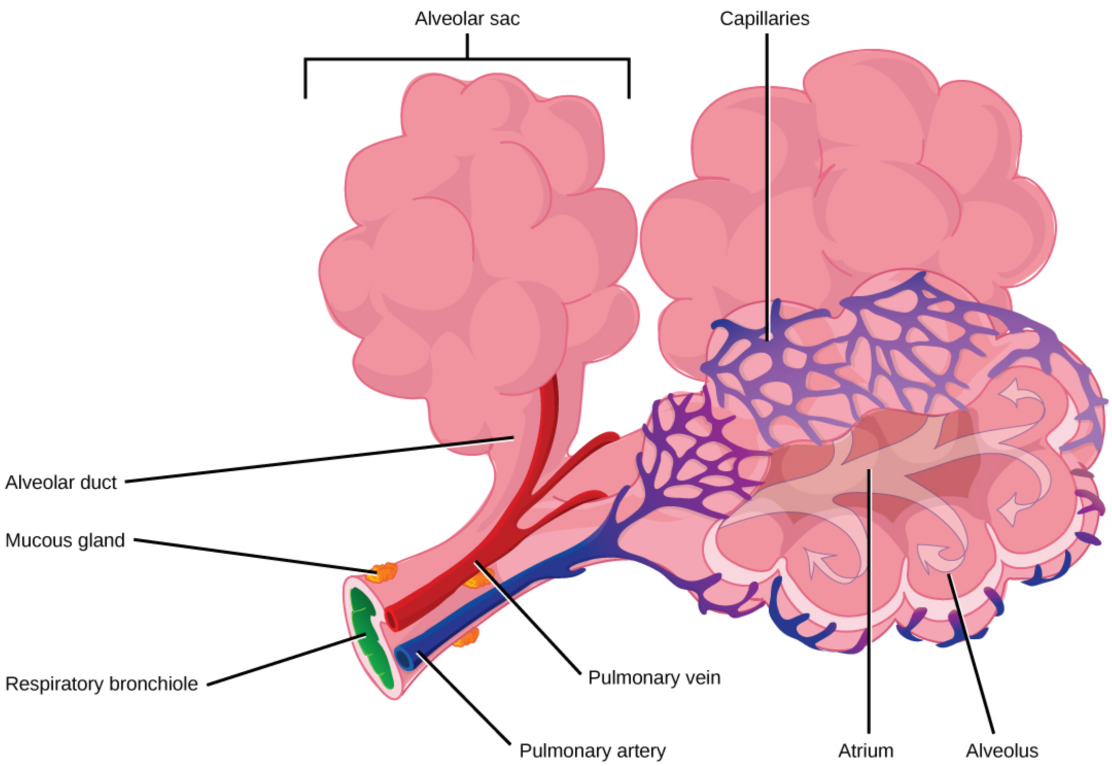


The trachea bifurcates into the right and left bronchi in the lungs. The right lung is made of three lobes and is larger. To accommodate the heart, the left lung is smaller and has only two lobes.

In the lungs, air is diverted into smaller and smaller passages, or **bronchi**. Air enters the lungs through the two **primary (main) bronchi** (singular: bronchus). Each bronchus divides into secondary bronchi, then into tertiary bronchi, which in turn divide, creating smaller and smaller diameter **bronchioles** as they split and spread through the lung. Like the trachea, the bronchi are made of cartilage and smooth muscle. At the bronchioles, the cartilage is replaced with elastic fibers. Bronchi are innervated by nerves of both the parasympathetic and sympathetic nervous systems that control muscle contraction (parasympathetic) or relaxation (sympathetic) in the bronchi and bronchioles, depending on the nervous system's cues. In humans, bronchioles with a diameter smaller than 0.5 mm are the **respiratory bronchioles**. They lack cartilage and therefore rely on inhaled air to support their shape. As the passageways decrease in diameter, the relative amount of smooth muscle increases.

The **terminal bronchioles** subdivide into microscopic branches called respiratory bronchioles. The respiratory bronchioles subdivide into several alveolar ducts. Numerous alveoli and alveolar sacs surround the alveolar ducts. The alveolar sacs resemble bunches of grapes tethered to the end of the bronchioles ([link]). In the acinar region, the **alveolar ducts** are attached to the end of each bronchiole. At the end of each duct are approximately 100 **alveolar sacs**, each containing 20 to 30 **alveoli** that are 200 to 300 microns in diameter. Gas exchange occurs only in alveoli. Alveoli are made of thin-walled parenchymal cells, typically one-cell thick, that look like tiny bubbles within the sacs. Alveoli are in direct contact with capillaries (one-cell thick) of the circulatory system. Such intimate contact ensures that oxygen will diffuse from alveoli into the blood and be distributed to the cells of the body. In addition, the carbon dioxide that was produced by cells as a waste product will diffuse from the blood into alveoli to be exhaled. The anatomical arrangement of capillaries and alveoli emphasizes the structural and functional relationship of the respiratory and circulatory systems. Because there are so many alveoli (~300 million per lung) within each alveolar sac and so many sacs at the end of each alveolar duct, the lungs have a sponge-like consistency. This organization produces a very large surface area that is available for gas exchange. The surface area of alveoli in the lungs is approximately 75 m^2. This large surface area,

combined with the thin-walled nature of the alveolar parenchymal cells, allows gases to easily diffuse across the cells.



Terminal bronchioles are connected by respiratory bronchioles to alveolar ducts and alveolar sacs. Each alveolar sac contains 20 to 30 spherical alveoli and has the appearance of a bunch of grapes. Air flows into the atrium of the alveolar sac, then circulates into alveoli where gas exchange occurs with the capillaries. Mucous glands secrete mucous into the airways, keeping them moist and flexible. (credit: modification of work by Mariana Ruiz Villareal)

Note:
Link to Learning

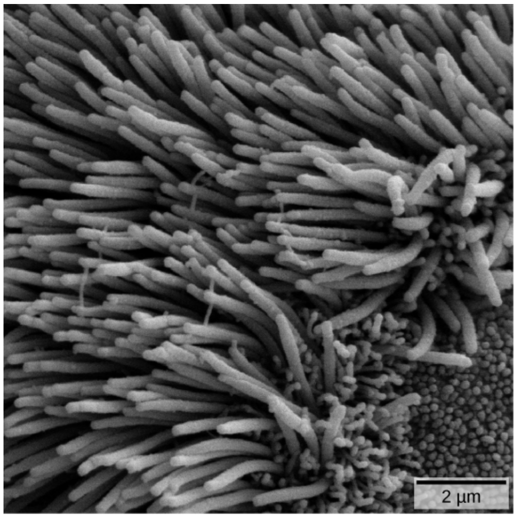
Watch the following video to review the respiratory system. https://www.openstaxcollege.org/l/lungs_pulmonary

Protective Mechanisms

The air that organisms breathe contains **particulate matter** such as dust, dirt, viral particles, and bacteria that can damage the lungs or trigger allergic immune responses. The respiratory system contains several protective mechanisms to avoid problems or tissue damage. In the nasal cavity, hairs and mucus trap small particles, viruses, bacteria, dust, and dirt to prevent their entry.

If particulates do make it beyond the nose, or enter through the mouth, the bronchi and bronchioles of the lungs also contain several protective devices. The lungs produce **mucus**—a sticky substance made of **mucin**, a complex glycoprotein, as well as salts and water—that traps particulates. The bronchi and bronchioles contain cilia, small hair-like projections that line the walls of the bronchi and bronchioles ([link]). These cilia beat in unison and move mucus and particles out of the bronchi and bronchioles back up to the throat where it is swallowed and eliminated via the esophagus.

In humans, for example, tar and other substances in cigarette smoke destroy or paralyze the cilia, making the removal of particles more difficult. In addition, smoking causes the lungs to produce more mucus, which the damaged cilia are not able to move. This causes a persistent cough, as the lungs try to rid themselves of particulate matter, and makes smokers more susceptible to respiratory ailments.



The bronchi and bronchioles contain cilia that help move mucus and other particles out of the lungs. (credit: Louisa Howard, modification of work by Dartmouth Electron Microscope Facility)

Section Summary

Animal respiratory systems are designed to facilitate gas exchange. In mammals, air is warmed and humidified in the nasal cavity. Air then travels down the pharynx, through the trachea, and into the lungs. In the lungs, air passes through the branching bronchi, reaching the respiratory bronchioles, which house the first site of gas exchange. The respiratory bronchioles open into the alveolar ducts, alveolar sacs, and alveoli. Because there are so many alveoli and alveolar sacs in the lung, the surface area for gas exchange is very large. Several protective mechanisms are in place to prevent damage or infection. These include the hair and mucus in the nasal cavity that trap dust, dirt, and other particulate matter before they can enter

the system. In the lungs, particles are trapped in a mucus layer and transported via cilia up to the esophageal opening at the top of the trachea to be swallowed.

Exercise:

Problem:

[link] Which of the following statements about the mammalian respiratory system is false?

a. When we breathe in, air travels from the pharynx to the trachea.
b. The bronchioles branch into bronchi.
c. Alveolar ducts connect to alveolar sacs.
d. Gas exchange between the lung and blood takes place in the alveolus.

Solution:

[link] B

Review Questions

Exercise:

Problem: The respiratory system _________.

a. provides body tissues with oxygen
b. provides body tissues with oxygen and carbon dioxide
c. establishes how many breaths are taken per minute
d. provides the body with carbon dioxide

Solution:

A

Exercise:

Problem:

Air is warmed and humidified in the nasal passages. This helps to ________.

a. ward off infection
b. decrease sensitivity during breathing
c. prevent damage to the lungs
d. all of the above

Solution:

C

Exercise:

Problem: Which is the order of airflow during inhalation?

a. nasal cavity, trachea, larynx, bronchi, bronchioles, alveoli
b. nasal cavity, larynx, trachea, bronchi, bronchioles, alveoli
c. nasal cavity, larynx, trachea, bronchioles, bronchi, alveoli
d. nasal cavity, trachea, larynx, bronchi, bronchioles, alveoli

---

Solution:

B

Free Response

Exercise:

Problem:

Describe the function of these terms and describe where they are located: main bronchus, trachea, alveoli, and acinus.

Solution:

The main bronchus is the conduit in the lung that funnels air to the airways where gas exchange occurs. The main bronchus attaches the lungs to the very end of the trachea where it bifurcates. The trachea is the cartilaginous structure that extends from the pharynx to the primary bronchi. It serves to funnel air to the lungs. The alveoli are the sites of gas exchange; they are located at the terminal regions of the lung and are attached to the respiratory bronchioles. The acinus is the structure in the lung where gas exchange occurs.

Exercise:

Problem: How does the structure of alveoli maximize gas exchange?

Solution:

The sac-like structure of the alveoli increases their surface area. In addition, the alveoli are made of thin-walled parenchymal cells. These features allow gases to easily diffuse across the cells.

Glossary

alveolar duct
: duct that extends from the terminal bronchiole to the alveolar sac

alveolar sac
: structure consisting of two or more alveoli that share a common opening

alveolus
: (plural: alveoli) (also, air sac) terminal region of the lung where gas exchange occurs

bronchus
: (plural: bronchi) smaller branch of cartilaginous tissue that stems off of the trachea; air is funneled through the bronchi to the region where gas

exchange occurs in alveoli

bronchiole
airway that extends from the main tertiary bronchi to the alveolar sac

diaphragm
domed-shaped skeletal muscle located under lungs that separates the thoracic cavity from the abdominal cavity

larynx
voice box, a short passageway connecting the pharynx and the trachea

mucin
complex glycoprotein found in mucus

mucus
sticky protein-containing fluid secretion in the lung that traps particulate matter to be expelled from the body

nasal cavity
opening of the respiratory system to the outside environment

particulate matter
small particle such as dust, dirt, viral particles, and bacteria that are in the air

pharynx
throat; a tube that starts in the internal nares and runs partway down the neck, where it opens into the esophagus and the larynx

primary bronchus
(also, main bronchus) region of the airway within the lung that attaches to the trachea and bifurcates to each lung where it branches into secondary bronchi

respiratory bronchiole
terminal portion of the bronchiole tree that is attached to the terminal bronchioles and alveoli ducts, alveolar sacs, and alveoli

terminal bronchiole
: region of bronchiole that attaches to the respiratory bronchioles

trachea
: cartilaginous tube that transports air from the larynx to the primary bronchi

Derived copy of Gas Exchange across Respiratory Surfaces
By the end of this section, you will be able to:

- Name and describe lung volumes and capacities
- Understand how gas pressure influences how gases move into and out of the body

The structure of the lung maximizes its surface area to increase gas diffusion. Because of the enormous number of alveoli (approximately 300 million in each human lung), the surface area of the lung is very large (75 m^2). Having such a large surface area increases the amount of gas that can diffuse into and out of the lungs.

Basic Principles of Gas Exchange

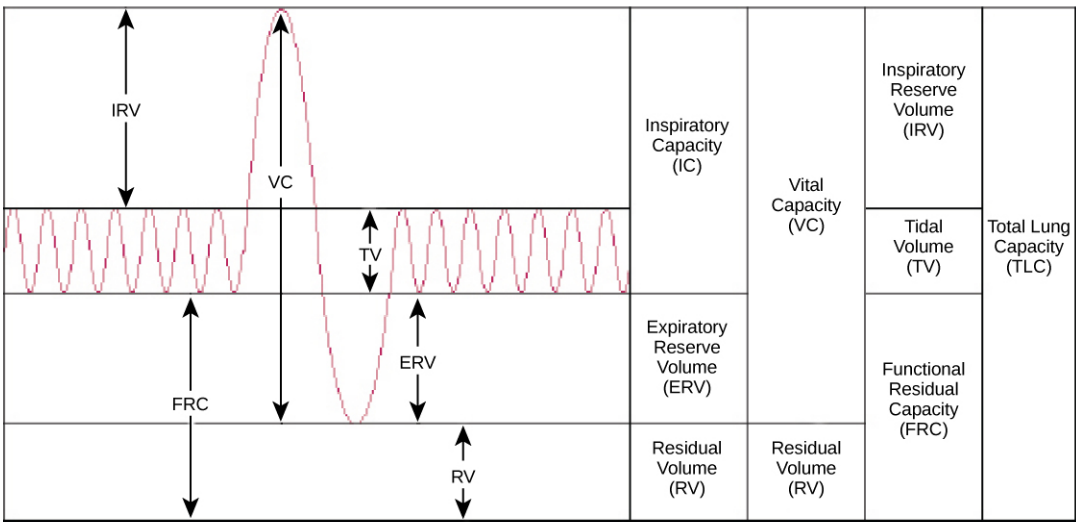
Gas exchange during respiration occurs primarily through diffusion. Diffusion is a process in which transport is driven by a concentration gradient. Gas molecules move from a region of high concentration to a region of low concentration. Blood that is low in oxygen concentration and high in carbon dioxide concentration undergoes gas exchange with air in the lungs. The air in the lungs has a higher concentration of oxygen than that of oxygen-depleted blood and a lower concentration of carbon dioxide. This concentration gradient allows for gas exchange during respiration.

Partial pressure is a measure of the concentration of the individual components in a mixture of gases. The total pressure exerted by the mixture is the sum of the partial pressures of the components in the mixture. The rate of diffusion of a gas is proportional to its partial pressure within the total gas mixture. This concept is discussed further in detail below.

Lung Volumes and Capacities

Different animals have different lung capacities based on their activities. Cheetahs have evolved a much higher lung capacity than humans; it helps provide oxygen to all the muscles in the body and allows them to run very fast. Elephants also have a high lung capacity. In this case, it is not because they run fast but because they have a large body and must be able to take up oxygen in accordance with their body size.

Human lung size is determined by genetics, gender, and height. At maximal capacity, an average lung can hold almost six liters of air, but lungs do not usually operate at maximal capacity. Air in the lungs is measured in terms of **lung volumes** and **lung capacities** ([link] and [link]). Volume measures the amount of air for one function (such as inhalation or exhalation). Capacity is any two or more volumes (for example, how much can be inhaled from the end of a maximal exhalation).



Human lung volumes and capacities are shown. The total lung capacity of the adult male is six liters. Tidal volume is the volume of air inhaled in a single, normal breath. Inspiratory capacity is the amount of air taken in during a deep breath, and residual volume is the amount of air left in the lungs after forceful respiration.

Lung Volumes and Capacities (Avg Adult Male)

Lung Volumes and Capacities (Avg Adult Male)

Volume/Capacity	**Definition**	**Volume (liters)**	**Equations**
Volume/Capacity	**Definition**	**Volume (liters)**	**Equations**
Tidal volume (TV)	Amount of air inhaled during a normal breath	0.5	-
Expiratory reserve volume (ERV)	Amount of air that can be exhaled after a normal exhalation	1.2	-
Inspiratory reserve volume (IRV)	Amount of air that can be further inhaled after a normal inhalation	3.1	-
Residual volume (RV)	Air left in the lungs after a forced exhalation	1.2	-

Lung Volumes and Capacities (Avg Adult Male)			
Volume/Capacity	**Definition**	**Volume (liters)**	**Equations**
Vital capacity (VC)	Maximum amount of air that can be moved in or out of the lungs in a single respiratory cycle	4.8	ERV+TV+IRV
Inspiratory capacity (IC)	Volume of air that can be inhaled in addition to a normal exhalation	3.6	TV+IRV
Functional residual capacity (FRC)	Volume of air remaining after a normal exhalation	2.4	ERV+RV

Lung Volumes and Capacities (Avg Adult Male)

Volume/Capacity	Definition	Volume (liters)	Equations
Total lung capacity (TLC)	Total volume of air in the lungs after a maximal inspiration	6.0	RV+ERV+TV+IRV
Forced expiratory volume (FEV1)	How much air can be forced out of the lungs over a specific time period, usually one second	~4.1 to 5.5	-

Lung volumes are measured by a technique called **spirometry**. An important measurement taken during spirometry is the **forced expiratory volume (FEV)**, which measures how much air can be forced out of the lung over a specific period, usually one second (FEV1). In addition, the forced vital capacity (FVC), which is the total amount of air that can be forcibly exhaled, is measured. The ratio of these values (**FEV1/FVC ratio**) is used to diagnose lung diseases including asthma, emphysema, and fibrosis. If the FEV1/FVC ratio is high, the lungs are not compliant (meaning they are stiff and unable to bend properly), and the patient most likely has lung fibrosis. Patients exhale most of the lung volume very quickly. Conversely, when the FEV1/FVC ratio is low, there is resistance in the lung that is characteristic of asthma. In this

instance, it is hard for the patient to get the air out of his or her lungs, and it takes a long time to reach the maximal exhalation volume. In either case, breathing is difficult and complications arise.

Note:
Career Connection
Respiratory Therapist
Respiratory therapists or respiratory practitioners evaluate and treat patients with lung and cardiovascular diseases. They work as part of a medical team to develop treatment plans for patients. Respiratory therapists may treat premature babies with underdeveloped lungs, patients with chronic conditions such as asthma, or older patients suffering from lung disease such as emphysema and chronic obstructive pulmonary disease (COPD). They may operate advanced equipment such as compressed gas delivery systems, ventilators, blood gas analyzers, and resuscitators. Specialized programs to become a respiratory therapist generally lead to a bachelor's degree with a respiratory therapist specialty. Because of a growing aging population, career opportunities as a respiratory therapist are expected to remain strong.

Gas Pressure and Respiration

The respiratory process can be better understood by examining the properties of gases. Gases move freely, but gas particles are constantly hitting the walls of their vessel, thereby producing gas pressure.

Air is a mixture of gases, primarily nitrogen (N_2; 78.6 percent), oxygen (O_2; 20.9 percent), water vapor (H_2O; 0.5 percent), and carbon dioxide (CO_2; 0.04 percent). Each gas component of that mixture exerts a pressure. The pressure for an individual gas in the mixture is the partial pressure of that gas. Approximately 21 percent of atmospheric gas is oxygen. Carbon dioxide, however, is found in relatively small amounts, 0.04 percent. The partial pressure for oxygen is much greater than that of carbon dioxide. The partial pressure of any gas can be calculated by:
Equation:

$$P = (P_{atm}) \times (\text{percent content in mixture}).$$

P_{atm}, the atmospheric pressure, is the sum of all of the partial pressures of the atmospheric gases added together,

Equation:

$$P_{atm} = P_{N_2} + P_{O_2} + P_{H_2O} + P_{CO_2} = 760 \text{ mm Hg}$$

× (percent content in mixture).

The pressure of the atmosphere at sea level is 760 mm Hg. Therefore, the partial pressure of oxygen is:

Equation:

$$P_{O_2} = (760 \text{ mm Hg})\ (0.21) = 160 \text{ mm Hg}$$

and for carbon dioxide:

Equation:

$$P_{CO_2} = (760 \text{ mm Hg})\ (0.0004) = 0.3 \text{ mm Hg}.$$

At high altitudes, P_{atm} decreases but concentration does not change; the partial pressure decrease is due to the reduction in P_{atm}.

When the air mixture reaches the lung, it has been humidified. The pressure of the water vapor in the lung does not change the pressure of the air, but it must be included in the partial pressure equation. For this calculation, the water pressure (47 mm Hg) is subtracted from the atmospheric pressure:

Equation:

$$760 \text{ mm Hg} - 47 \text{ mm Hg} = 713 \text{ mm Hg}$$

and the partial pressure of oxygen is:

Equation:

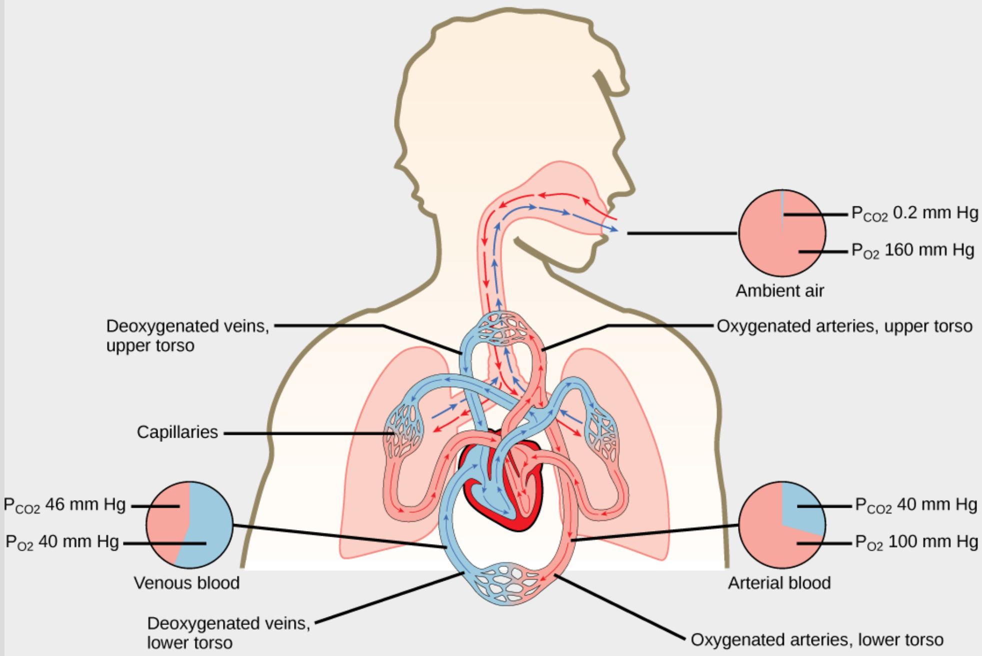
$$(760 \text{ mm Hg} - 47 \text{ mm Hg}) \times 0.21 = 150 \text{ mm Hg}.$$

These pressures determine the gas exchange, or the flow of gas, in the system. Oxygen and carbon dioxide will flow according to their pressure gradient from high to low. Therefore, understanding the partial pressure of each gas will aid in understanding how gases move in the respiratory system.

Gas Exchange across the Alveoli

Oxygen and carbon dioxide move independently of each other; they diffuse down their own pressure gradients. As blood leaves the lungs through the pulmonary veins, the **venous** P_{O_2}= 100 mm Hg, whereas the **venous** P_{CO_2} = 40 mm Hg. As blood enters the systemic capillaries, the blood will lose oxygen and gain carbon dioxide because of the pressure difference of the tissues and blood. In systemic capillaries, P_{O_2}= 100 mm Hg, but in the tissue cells, P_{O_2}= 40 mm Hg. This pressure gradient drives the diffusion of oxygen out of the capillaries and into the tissue cells. At the same time, blood P_{CO_2}= 40 mm Hg and systemic tissue P_{CO_2}= 45 mm Hg. The pressure gradient drives CO_2 out of tissue cells and into the capillaries. The blood returning to the lungs through the pulmonary arteries has a venous P_{O_2}= 40 mm Hg and a P_{CO_2}= 45 mm Hg. The blood enters the lung capillaries where the process of exchanging gases between the capillaries and alveoli begins again ([link]).

Note:
Art Connection



The partial pressures of oxygen and carbon dioxide change as blood moves through the body.

Which of the following statements is false?

a. In the tissues, P_{O_2} drops as blood passes from the arteries to the veins, while P_{CO_2} increases.
b. Blood travels from the lungs to the heart to body tissues, then back to the heart, then the lungs.
c. Blood travels from the lungs to the heart to body tissues, then back to the lungs, then the heart.
d. P_{O_2} is higher in air than in the lungs.

In short, the change in partial pressure from the alveoli to the capillaries drives the oxygen into the tissues and the carbon dioxide into the blood from the tissues. The blood is then transported to the lungs where differences in pressure in the alveoli result in the movement of carbon dioxide out of the blood into the lungs, and oxygen into the blood.

Note:
Link to Learning

Watch this video to learn how to carry out spirometry.

Section Summary

The lungs can hold a large volume of air, but they are not usually filled to maximal capacity. Lung volume measurements include tidal volume, expiratory reserve volume, inspiratory reserve volume, and residual volume. The sum of these equals the total lung capacity. Gas movement into or out of the lungs is dependent on the pressure of the gas. Air is a mixture of gases; therefore, the partial pressure of each gas can be calculated to determine how the gas will flow in the lung. The difference between the partial pressure of the gas in the air drives oxygen into the tissues and carbon dioxide out of the body.

Art Connections

Exercise:

Problem: [link] Which of the following statements is false?

a. In the tissues, P_{O_2} drops as blood passes from the arteries to the veins, while P_{CO_2} increases.
b. Blood travels from the lungs to the heart to body tissues, then back to the heart, then the lungs.
c. Blood travels from the lungs to the heart to body tissues, then back to the lungs, then the heart.

d. P_{O_2} is higher in air than in the lungs.

Solution:

[link] C

Review Questions

Exercise:

Problem: The inspiratory reserve volume measures the ________.

a. amount of air remaining in the lung after a maximal exhalation
b. amount of air that the lung holds
c. amount of air the can be further exhaled after a normal breath
d. amount of air that can be further inhaled after a normal breath

---

Solution:

D

Exercise:

Problem:

Of the following, which does not explain why the partial pressure of oxygen is lower in the lung than in the external air?

a. Air in the lung is humidified; therefore, water vapor pressure alters the pressure.
b. Carbon dioxide mixes with oxygen.
c. Oxygen is moved into the blood and is headed to the tissues.
d. Lungs exert a pressure on the air to reduce the oxygen pressure.

Solution:

D

Exercise:

Problem:

The total lung capacity is calculated using which of the following formulas?

a. residual volume + tidal volume + inspiratory reserve volume
b. residual volume + expiratory reserve volume + inspiratory reserve volume
c. expiratory reserve volume + tidal volume + inspiratory reserve volume
d. residual volume + expiratory reserve volume + tidal volume + inspiratory reserve volume

---

Solution:

D

Glossary

alveolar P_{O_2}
: partial pressure of oxygen in the alveoli (usually around 100 mmHg)

expiratory reserve volume (ERV)
: amount of additional air that can be exhaled after a normal exhalation

FEV1/FVC ratio
: ratio of how much air can be forced out of the lung in one second to the total amount that is forced out of the lung; a measurement of lung function that can be used to detect disease states

forced expiratory volume (FEV)
: (also, forced vital capacity) measure of how much air can be forced out of the lung from maximal inspiration over a specific amount of time

functional residual capacity (FRC)
expiratory reserve volume plus residual volume

inspiratory capacity (IC)
tidal volume plus inspiratory reserve volume

inspiratory reserve volume (IRV)
amount of additional air that can be inspired after a normal inhalation

lung capacity
measurement of two or more lung volumes (how much air can be inhaled from the end of an expiration to maximal capacity)

lung volume
measurement of air for one lung function (normal inhalation or exhalation)

partial pressure
amount of pressure exerted by one gas within a mixture of gases

residual volume (RV)
amount of air remaining in the lung after a maximal expiration

respiratory quotient (RQ)
ratio of carbon dioxide production to each oxygen molecule consumed

spirometry
method to measure lung volumes and to diagnose lung diseases

tidal volume (TV)
amount of air that is inspired and expired during normal breathing

total lung capacity (TLC)
sum of the residual volume, expiratory reserve volume, tidal volume, and inspiratory reserve volume

venous P_{CO_2}
partial pressure of carbon dioxide in the veins (40 mm Hg in the pulmonary veins)

venous P_{O_2}

partial pressure of oxygen in the veins (100 mm Hg in the pulmonary veins)

vital capacity (VC)

sum of the expiratory reserve volume, tidal volume, and inspiratory reserve volume

Breathing
By the end of this section, you will be able to:

- Describe how the structures of the lungs and thoracic cavity control the mechanics of breathing
- Explain the importance of compliance and resistance in the lungs
- Discuss problems that may arise due to a V/Q mismatch

Mammalian lungs are located in the thoracic cavity where they are surrounded and protected by the rib cage, intercostal muscles, and bound by the chest wall. The bottom of the lungs is contained by the diaphragm, a skeletal muscle that facilitates breathing. Breathing requires the coordination of the lungs, the chest wall, and most importantly, the diaphragm.

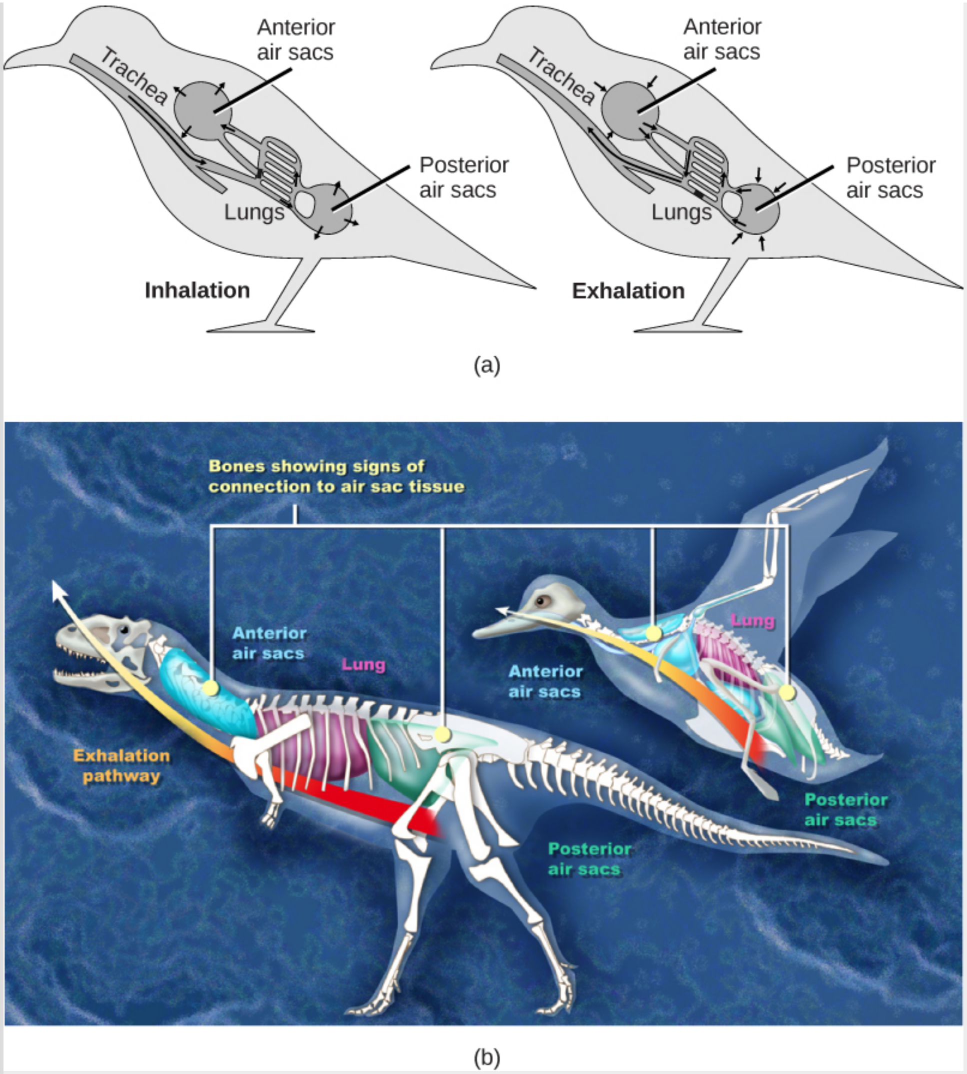
Types of Breathing

Amphibians have evolved multiple ways of breathing. Young amphibians, like tadpoles, use gills to breathe, and they don't leave the water. Some amphibians retain gills for life. As the tadpole grows, the gills disappear and lungs grow. These lungs are primitive and not as evolved as mammalian lungs. Adult amphibians are lacking or have a reduced diaphragm, so breathing via lungs is forced. The other means of breathing for amphibians is diffusion across the skin. To aid this diffusion, amphibian skin must remain moist.

Birds face a unique challenge with respect to breathing: They fly. Flying consumes a great amount of energy; therefore, birds require a lot of oxygen to aid their metabolic processes. Birds have evolved a respiratory system that supplies them with the oxygen needed to enable flying. Similar to mammals, birds have lungs, which are organs specialized for gas exchange. Oxygenated air, taken in during inhalation, diffuses across the surface of the lungs into the bloodstream, and carbon dioxide diffuses from the blood into the lungs and expelled during exhalation. The details of breathing between birds and mammals differ substantially.

In addition to lungs, birds have air sacs inside their body. Air flows in one direction from the posterior air sacs to the lungs and out of the anterior air sacs. The flow of air is in the opposite direction from blood flow, and gas exchange takes place much more efficiently. This type of breathing enables birds to obtain the requisite oxygen, even at higher altitudes where the oxygen concentration is low. This directionality of airflow requires two cycles of air intake and exhalation to completely get the air out of the lungs.

Note:
Evolution Connection
Avian Respiration
Birds have evolved a respiratory system that enables them to fly. Flying is a high-energy process and requires a lot of oxygen. Furthermore, many birds fly in high altitudes where the concentration of oxygen in low. How did birds evolve a respiratory system that is so unique?
Decades of research by paleontologists have shown that birds evolved from therapods, meat-eating dinosaurs ([link]). In fact, fossil evidence shows that meat-eating dinosaurs that lived more than 100 million years ago had a similar flow-through respiratory system with lungs and air sacs. *Archaeopteryx* and *Xiaotingia*, for example, were flying dinosaurs and are believed to be early precursors of birds.



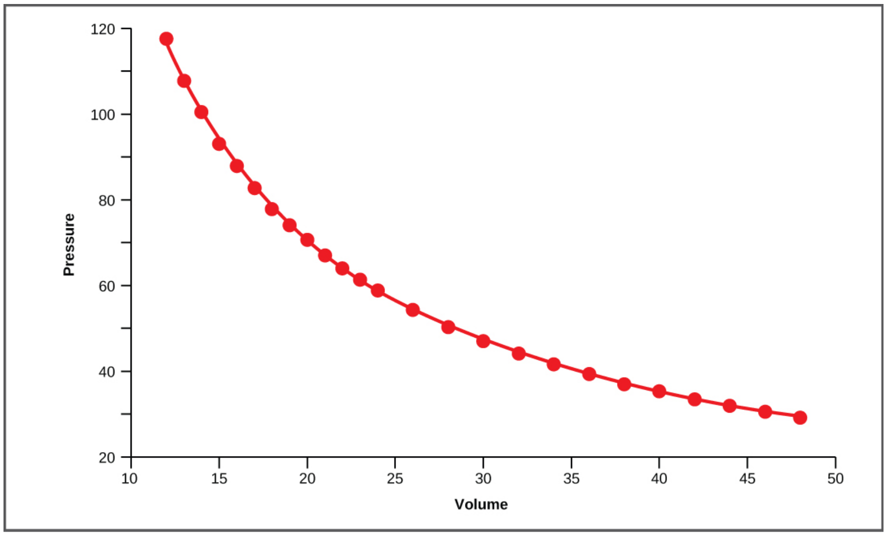
(a) Birds have a flow-through respiratory system in which air flows unidirectionally from the posterior sacs into the lungs, then into the anterior air sacs. The air sacs connect to openings in hollow bones. (b) Dinosaurs, from which birds descended, have similar hollow bones and are believed to have had a similar respiratory system. (credit b: modification of work by Zina Deretsky, National Science Foundation)

Most of us consider that dinosaurs are extinct. However, modern birds are descendants of avian dinosaurs. The respiratory system of modern birds has been evolving for hundreds of millions of years.

All mammals have lungs that are the main organs for breathing. Lung capacity has evolved to support the animal's activities. During inhalation, the lungs expand with air, and oxygen diffuses across the lung's surface and enters the bloodstream. During exhalation, the lungs expel air and lung volume decreases. In the next few sections, the process of human breathing will be explained.

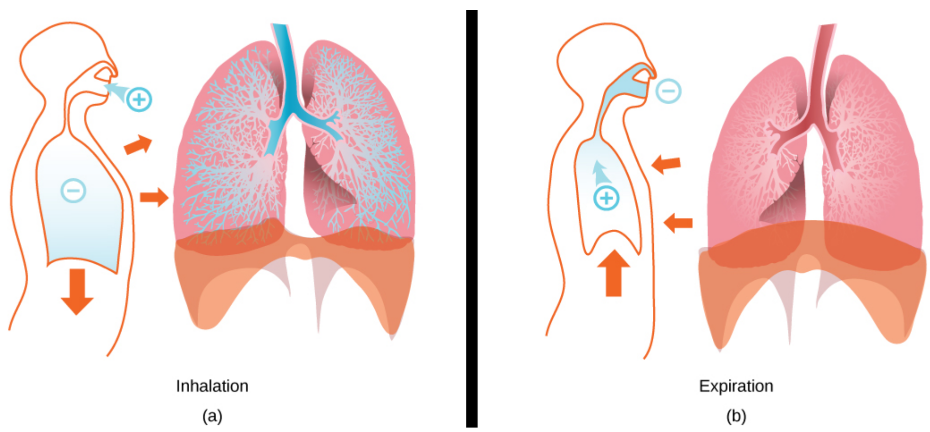
The Mechanics of Human Breathing

Boyle's Law is the gas law that states that in a closed space, pressure and volume are inversely related. As volume decreases, pressure increases and vice versa ([link]). The relationship between gas pressure and volume helps to explain the mechanics of breathing.



This graph shows data from Boyle's original 1662 experiment, which shows that pressure and volume are inversely related. No units are given as Boyle used arbitrary units in his experiments.

There is always a slightly negative pressure within the thoracic cavity, which aids in keeping the airways of the lungs open. During inhalation, volume increases as a result of contraction of the diaphragm, and pressure decreases (according to Boyle's Law). This decrease of pressure in the thoracic cavity relative to the environment makes the cavity less than the atmosphere ([link]**a**). Because of this drop in pressure, air rushes into the respiratory passages. To increase the volume of the lungs, the chest wall expands. This results from the contraction of the **intercostal muscles**, the muscles that are connected to the rib cage. Lung volume expands because the diaphragm contracts and the intercostals muscles contract, thus expanding the thoracic cavity. This increase in the volume of the thoracic cavity lowers pressure compared to the atmosphere, so air rushes into the lungs, thus increasing its volume. The resulting increase in volume is largely attributed to an increase in alveolar space, because the bronchioles and bronchi are stiff structures that do not change in size.

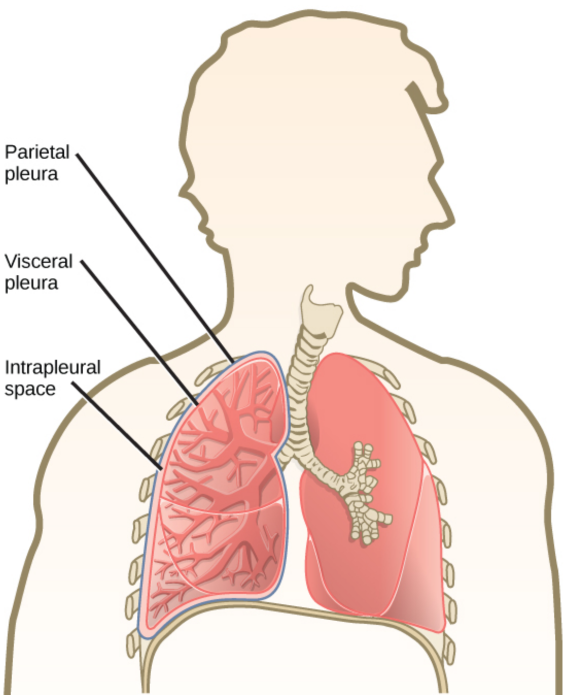


The lungs, chest wall, and diaphragm are all involved in respiration, both (a) inhalation and (b) expiration. (credit: modification of work by Mariana Ruiz Villareal)

The chest wall expands out and away from the lungs. The lungs are elastic; therefore, when air fills the lungs, the **elastic recoil** within the tissues of the lung exerts pressure back toward the interior of the lungs. These outward

and inward forces compete to inflate and deflate the lung with every breath. Upon exhalation, the lungs recoil to force the air out of the lungs, and the intercostal muscles relax, returning the chest wall back to its original position ([link]**b**). The diaphragm also relaxes and moves higher into the thoracic cavity. This increases the pressure within the thoracic cavity relative to the environment, and air rushes out of the lungs. The movement of air out of the lungs is a passive event. No muscles are contracting to expel the air.

Each lung is surrounded by an invaginated sac. The layer of tissue that covers the lung and dips into spaces is called the visceral **pleura**. A second layer of parietal pleura lines the interior of the thorax ([link]). The space between these layers, the **intrapleural space**, contains a small amount of fluid that protects the tissue and reduces the friction generated from rubbing the tissue layers together as the lungs contract and relax. **Pleurisy** results when these layers of tissue become inflamed; it is painful because the inflammation increases the pressure within the thoracic cavity and reduces the volume of the lung.



A tissue layer called pleura surrounds the lung and interior of the thoracic cavity. (credit: modification of work by NCI)

Note:
Link to Learning

View how Boyle’s Law is related to breathing and watch a video on Boyle’s Law.
https://www.openstaxcollege.org/l/boyle_breathing

The Work of Breathing

The number of breaths per minute is the **respiratory rate**. On average, under non-exertion conditions, the human respiratory rate is 12–15 breaths/minute. The respiratory rate contributes to the **alveolar ventilation**, or how much air moves into and out of the alveoli. Alveolar ventilation prevents carbon dioxide buildup in the alveoli. There are two ways to keep the alveolar ventilation constant: increase the respiratory rate while decreasing the tidal volume of air per breath (shallow breathing), or decrease the respiratory rate while increasing the tidal volume per breath. In either case, the ventilation remains the same, but the work done and type of

work needed are quite different. Both tidal volume and respiratory rate are closely regulated when oxygen demand increases.

There are two types of work conducted during respiration, flow-resistive and elastic work. **Flow-resistive** refers to the work of the alveoli and tissues in the lung, whereas **elastic work** refers to the work of the intercostal muscles, chest wall, and diaphragm. Increasing the respiration rate increases the flow-resistive work of the airways and decreases the elastic work of the muscles. Decreasing the respiratory rate reverses the type of work required.

Surfactant

The air-tissue/water interface of the alveoli has a high surface tension. This surface tension is similar to the surface tension of water at the liquid-air interface of a water droplet that results in the bonding of the water molecules together. **Surfactant** is a complex mixture of phospholipids and lipoproteins that works to reduce the surface tension that exists between the alveoli tissue and the air found within the alveoli. By lowering the surface tension of the alveolar fluid, it reduces the tendency of alveoli to collapse.

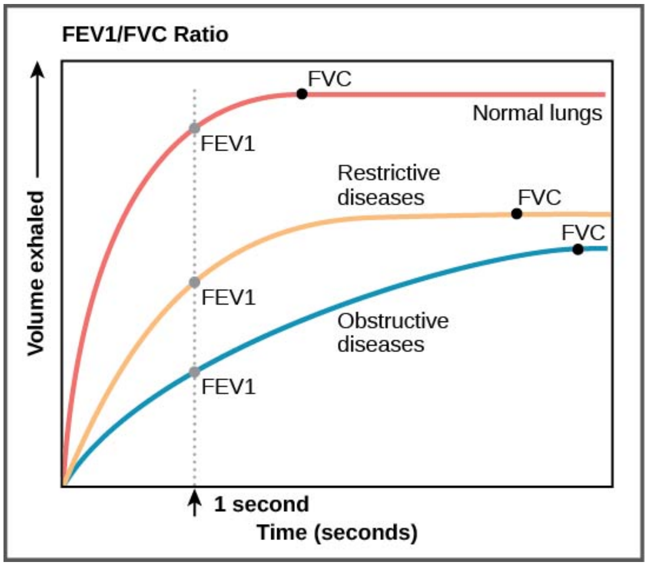
Surfactant works like a detergent to reduce the surface tension and allows for easier inflation of the airways. When a balloon is first inflated, it takes a large amount of effort to stretch the plastic and start to inflate the balloon. If a little bit of detergent was applied to the interior of the balloon, then the amount of effort or work needed to begin to inflate the balloon would decrease, and it would become much easier to start blowing up the balloon. This same principle applies to the airways. A small amount of surfactant to the airway tissues reduces the effort or work needed to inflate those airways. Babies born prematurely sometimes do not produce enough surfactant. As a result, they suffer from **respiratory distress syndrome**, because it requires more effort to inflate their lungs. Surfactant is also important for preventing collapse of small alveoli relative to large alveoli.

Lung Resistance and Compliance

Pulmonary diseases reduce the rate of gas exchange into and out of the lungs. Two main causes of decreased gas exchange are **compliance** (how elastic the lung is) and **resistance** (how much obstruction exists in the airways). A change in either can dramatically alter breathing and the ability to take in oxygen and release carbon dioxide.

Examples of **restrictive diseases** are respiratory distress syndrome and pulmonary fibrosis. In both diseases, the airways are less compliant and they are stiff or fibrotic. There is a decrease in compliance because the lung tissue cannot bend and move. In these types of restrictive diseases, the intrapleural pressure is more positive and the airways collapse upon exhalation, which traps air in the lungs. Forced or **functional vital capacity (FVC)**, which is the amount of air that can be forcibly exhaled after taking the deepest breath possible, is much lower than in normal patients, and the time it takes to exhale most of the air is greatly prolonged ([link]). A patient suffering from these diseases cannot exhale the normal amount of air.

Obstructive diseases and conditions include emphysema, asthma, and pulmonary edema. In emphysema, which mostly arises from smoking tobacco, the walls of the alveoli are destroyed, decreasing the surface area for gas exchange. The overall compliance of the lungs is increased, because as the alveolar walls are damaged, lung elastic recoil decreases due to a loss of elastic fibers, and more air is trapped in the lungs at the end of exhalation. Asthma is a disease in which inflammation is triggered by environmental factors. Inflammation obstructs the airways. The obstruction may be due to edema (fluid accumulation), smooth muscle spasms in the walls of the bronchioles, increased mucus secretion, damage to the epithelia of the airways, or a combination of these events. Those with asthma or edema experience increased occlusion from increased inflammation of the airways. This tends to block the airways, preventing the proper movement of gases ([link]). Those with obstructive diseases have large volumes of air trapped after exhalation and breathe at a very high lung volume to compensate for the lack of airway recruitment.



The ratio of FEV1 (the amount of air that can be forcibly exhaled in one second after taking a deep breath) to FVC (the total amount of air that can be forcibly exhaled) can be used to diagnose whether a person has restrictive or obstructive lung disease. In restrictive lung disease, FVC is reduced but airways are not obstructed, so the person is able to expel air reasonably fast. In obstructive lung disease, airway obstruction results in slow exhalation as well as reduced FVC. Thus, the FEV1/FVC ratio is lower in persons with obstructive lung disease (less than 69 percent) than in persons with restrictive disease (88 to 90 percent).

Dead Space: V/Q Mismatch

Pulmonary circulation pressure is very low compared to that of the systemic circulation. It is also independent of cardiac output. This is because of a phenomenon called **recruitment**, which is the process of opening airways that normally remain closed when cardiac output increases. As cardiac output increases, the number of capillaries and arteries that are perfused (filled with blood) increases. These capillaries and arteries are not always in use but are ready if needed. At times, however, there is a mismatch between the amount of air (ventilation, V) and the amount of blood (perfusion, Q) in the lungs. This is referred to as **ventilation/perfusion (V/Q) mismatch**.

There are two types of V/Q mismatch. Both produce **dead space**, regions of broken down or blocked lung tissue. Dead spaces can severely impact breathing, because they reduce the surface area available for gas diffusion. As a result, the amount of oxygen in the blood decreases, whereas the carbon dioxide level increases. Dead space is created when no ventilation and/or perfusion takes place. **Anatomical dead space** or anatomical shunt, arises from an anatomical failure, while **physiological dead space** or physiological shunt, arises from a functional impairment of the lung or arteries.

An example of an anatomical shunt is the effect of gravity on the lungs. The lung is particularly susceptible to changes in the magnitude and direction of gravitational forces. When someone is standing or sitting upright, the pleural pressure gradient leads to increased ventilation further down in the lung. As a result, the intrapleural pressure is more negative at the base of the lung than at the top, and more air fills the bottom of the lung than the top. Likewise, it takes less energy to pump blood to the bottom of the lung than to the top when in a prone position. Perfusion of the lung is not uniform while standing or sitting. This is a result of hydrostatic forces combined with the effect of airway pressure. An anatomical shunt develops because the ventilation of the airways does not match the perfusion of the arteries surrounding those airways. As a result, the rate of gas exchange is reduced. Note that this does not occur when lying down, because in this position, gravity does not preferentially pull the bottom of the lung down.

A physiological shunt can develop if there is infection or edema in the lung that obstructs an area. This will decrease ventilation but not affect perfusion; therefore, the V/Q ratio changes and gas exchange is affected.

The lung can compensate for these mismatches in ventilation and perfusion. If ventilation is greater than perfusion, the arterioles dilate and the bronchioles constrict. This increases perfusion and reduces ventilation. Likewise, if ventilation is less than perfusion, the arterioles constrict and the bronchioles dilate to correct the imbalance.

Note:
Link to Learning

View the mechanics of breathing.
https://www.openstaxcollege.org/l/breathing

Section Summary

The structure of the lungs and thoracic cavity control the mechanics of breathing. Upon inspiration, the diaphragm contracts and lowers. The intercostal muscles contract and expand the chest wall outward. The intrapleural pressure drops, the lungs expand, and air is drawn into the airways. When exhaling, the intercostal muscles and diaphragm relax, returning the intrapleural pressure back to the resting state. The lungs recoil and airways close. The air passively exits the lung. There is high surface tension at the air-airway interface in the lung. Surfactant, a mixture of phospholipids and lipoproteins, acts like a detergent in the airways to reduce surface tension and allow for opening of the alveoli.

Breathing and gas exchange are both altered by changes in the compliance and resistance of the lung. If the compliance of the lung decreases, as occurs in restrictive diseases like fibrosis, the airways stiffen and collapse upon exhalation. Air becomes trapped in the lungs, making breathing more difficult. If resistance increases, as happens with asthma or emphysema, the airways become obstructed, trapping air in the lungs and causing breathing to become difficult. Alterations in the ventilation of the airways or perfusion of the arteries can affect gas exchange. These changes in ventilation and perfusion, called V/Q mismatch, can arise from anatomical or physiological changes.

Review Questions

Exercise:

Problem: How would paralysis of the diaphragm alter inspiration?

a. It would prevent contraction of the intercostal muscles.
b. It would prevent inhalation because the intrapleural pressure would not change.
c. It would decrease the intrapleural pressure and allow more air to enter the lungs.
d. It would slow expiration because the lung would not relax.

Solution:

B

Exercise:

Problem: Restrictive airway diseases _________.

a. increase the compliance of the lung
b. decrease the compliance of the lung
c. increase the lung volume
d. decrease the work of breathing

Solution:

B

Exercise:

Problem: Alveolar ventilation remains constant when ________.

a. the respiratory rate is increased while the volume of air per breath is decreased
b. the respiratory rate and the volume of air per breath are increased
c. the respiratory rate is decreased while increasing the volume per breath
d. both a and c

Solution:

D

Free Response

Exercise:

Problem:

How would increased airway resistance affect intrapleural pressure during inhalation?

Solution:

Increased airway resistance increases the volume and pressure in the lung; therefore, the intrapleural pressure would be less negative and breathing would be more difficult.

Exercise:

Problem:

Explain how a puncture to the thoracic cavity (from a knife wound, for instance) could alter the ability to inhale.

Solution:

A puncture to the thoracic cavity would equalize the pressure inside the thoracic cavity to the outside environment. For the lung to function properly, the intrapleural pressure must be negative. This is caused by the contraction of the diaphragm pulling the lungs down and drawing air into the lungs.

Exercise:

Problem:

When someone is standing, gravity stretches the bottom of the lung down toward the floor to a greater extent than the top of the lung. What implication could this have on the flow of air in the lungs? Where does gas exchange occur in the lungs?

Solution:

The lung is particularly susceptible to changes in the magnitude and direction of gravitational forces. When someone is standing or sitting upright, the pleural pressure gradient leads to increased ventilation further down in the lung.

Glossary

alveolar ventilation
: how much air is in the alveoli

anatomical dead space
: (also, anatomical shunt) region of the lung that lacks proper ventilation/perfusion due to an anatomical block

compliance
measurement of the elasticity of the lung

dead space
area in the lung that lacks proper ventilation or perfusion

elastic recoil
property of the lung that drives the lung tissue inward

elastic work
work conducted by the intercostal muscles, chest wall, and diaphragm

flow-resistive
work of breathing performed by the alveoli and tissues in the lung

functional vital capacity (FVC)
amount of air that can be forcibly exhaled after taking the deepest breath possible

intercostal muscle
muscle connected to the rib cage that contracts upon inspiration

intrapleural space
space between the layers of pleura

obstructive disease
disease (such as emphysema and asthma) that arises from obstruction of the airways; compliance increases in these diseases

physiological dead space
(also, physiological shunt) region of the lung that lacks proper ventilation/perfusion due to a physiological change in the lung (like inflammation or edema)

pleura
tissue layer that surrounds the lungs and lines the interior of the thoracic cavity

pleurisy

painful inflammation of the pleural tissue layers

recruitment
process of opening airways that normally remain closed when the cardiac output increases

resistance
measurement of lung obstruction

respiratory distress syndrome
disease that arises from a deficient amount of surfactant

respiratory rate
number of breaths per minute

restrictive disease
disease that results from a restriction and decreased compliance of the alveoli; respiratory distress syndrome and pulmonary fibrosis are examples

surfactant
detergent-like liquid in the airways that lowers the surface tension of the alveoli to allow for expansion

ventilation/perfusion (V/Q) mismatch
region of the lung that lacks proper alveolar ventilation (V) and/or arterial perfusion (Q)

Transport of Gases in Human Bodily Fluids
By the end of this section, you will be able to:

- Describe how oxygen is bound to hemoglobin and transported to body tissues
- Explain how carbon dioxide is transported from body tissues to the lungs

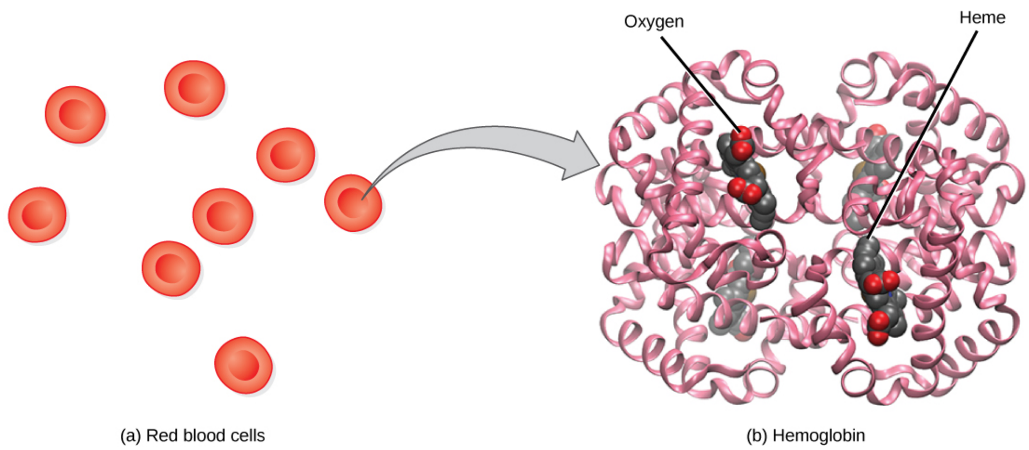
Once the oxygen diffuses across the alveoli, it enters the bloodstream and is transported to the tissues where it is unloaded, and carbon dioxide diffuses out of the blood and into the alveoli to be expelled from the body. Although gas exchange is a continuous process, the oxygen and carbon dioxide are transported by different mechanisms.

Transport of Oxygen in the Blood

Although oxygen dissolves in blood, only a small amount of oxygen is transported this way. Only 1.5 percent of oxygen in the blood is dissolved directly into the blood itself. Most oxygen—98.5 percent—is bound to a protein called hemoglobin and carried to the tissues.

Hemoglobin

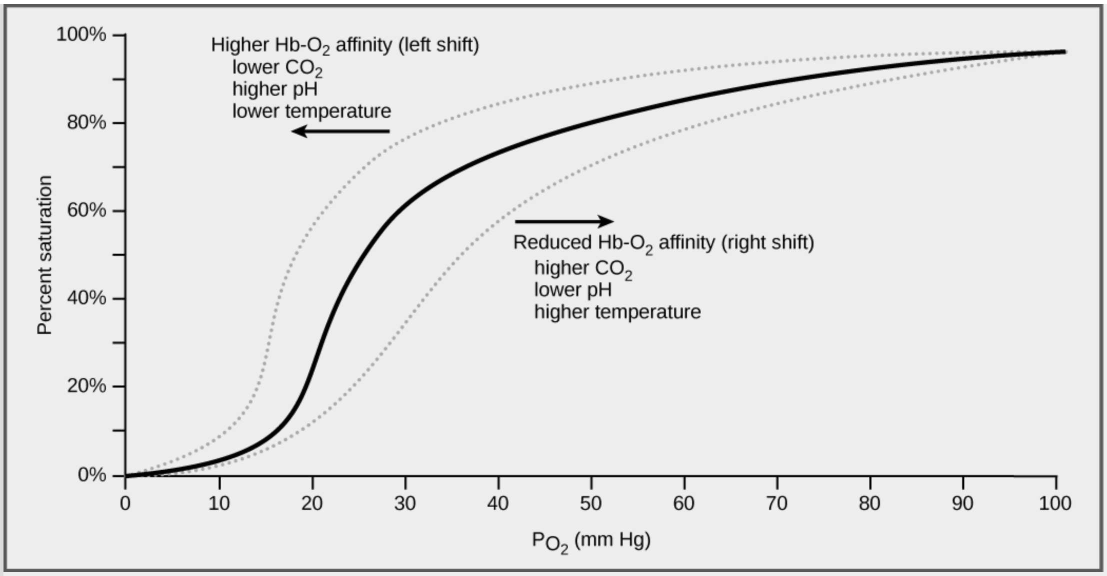
Hemoglobin, or Hb, is a protein molecule found in red blood cells (erythrocytes) made of four subunits: two alpha subunits and two beta subunits ([link]). Each subunit surrounds a central **heme group** that contains iron and binds one oxygen molecule, allowing each hemoglobin molecule to bind four oxygen molecules. Molecules with more oxygen bound to the heme groups are brighter red. As a result, oxygenated arterial blood where the Hb is carrying four oxygen molecules is bright red, while venous blood that is deoxygenated is darker red.



The protein inside (a) red blood cells that carries oxygen to cells and carbon dioxide to the lungs is (b) hemoglobin. Hemoglobin is made up of four symmetrical subunits and four heme groups. Iron associated with the heme binds oxygen. It is the iron in hemoglobin that gives blood its red color.

It is easier to bind a second and third oxygen molecule to Hb than the first molecule. This is because the hemoglobin molecule changes its shape, or conformation, as oxygen binds. The fourth oxygen is then more difficult to bind. The binding of oxygen to hemoglobin can be plotted as a function of the partial pressure of oxygen in the blood (x-axis) versus the relative Hb-oxygen saturation (y-axis). The resulting graph—an **oxygen dissociation curve**—is sigmoidal, or S-shaped ([link]). As the partial pressure of oxygen increases, the hemoglobin becomes increasingly saturated with oxygen.

Note:
Art Connection



The oxygen dissociation curve demonstrates that, as the partial pressure of oxygen increases, more oxygen binds hemoglobin. However, the affinity of hemoglobin for oxygen may shift to the left or the right depending on environmental conditions.

The kidneys are responsible for removing excess H+ ions from the blood. If the kidneys fail, what would happen to blood pH and to hemoglobin affinity for oxygen?

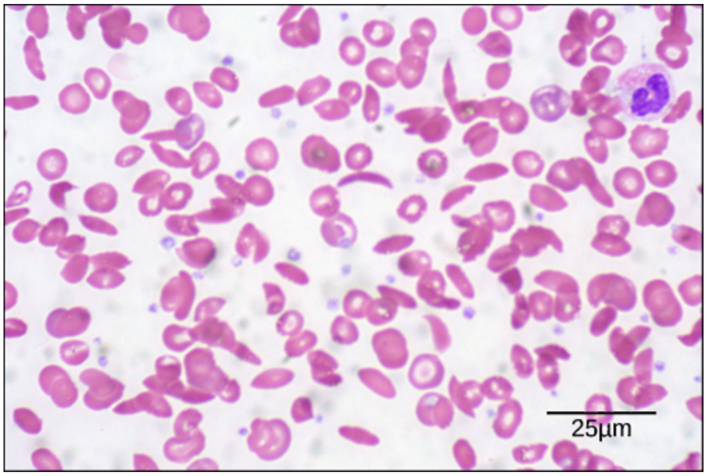
Factors That Affect Oxygen Binding

The **oxygen-carrying capacity** of hemoglobin determines how much oxygen is carried in the blood. In addition to P_{O_2}, other environmental factors and diseases can affect oxygen carrying capacity and delivery.

Carbon dioxide levels, blood pH, and body temperature affect oxygen-carrying capacity ([link]). When carbon dioxide is in the blood, it reacts with water to form bicarbonate $\left(HCO_3^-\right)$ and hydrogen ions (H^+). As the level of carbon dioxide in the blood increases, more H^+ is produced and the pH decreases. This increase in carbon dioxide and subsequent decrease in pH reduce the affinity of hemoglobin for oxygen. The oxygen dissociates

from the Hb molecule, shifting the oxygen dissociation curve to the right. Therefore, more oxygen is needed to reach the same hemoglobin saturation level as when the pH was higher. A similar shift in the curve also results from an increase in body temperature. Increased temperature, such as from increased activity of skeletal muscle, causes the affinity of hemoglobin for oxygen to be reduced.

Diseases like sickle cell anemia and thalassemia decrease the blood's ability to deliver oxygen to tissues and its oxygen-carrying capacity. In **sickle cell anemia**, the shape of the red blood cell is crescent-shaped, elongated, and stiffened, reducing its ability to deliver oxygen ([link]). In this form, red blood cells cannot pass through the capillaries. This is painful when it occurs. **Thalassemia** is a rare genetic disease caused by a defect in either the alpha or the beta subunit of Hb. Patients with thalassemia produce a high number of red blood cells, but these cells have lower-than-normal levels of hemoglobin. Therefore, the oxygen-carrying capacity is diminished.



Individuals with sickle cell anemia have crescent-shaped red blood cells. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

Transport of Carbon Dioxide in the Blood

Carbon dioxide molecules are transported in the blood from body tissues to the lungs by one of three methods: dissolution directly into the blood, binding to hemoglobin, or carried as a bicarbonate ion. Several properties of carbon dioxide in the blood affect its transport. First, carbon dioxide is more soluble in blood than oxygen. About 5 to 7 percent of all carbon dioxide is dissolved in the plasma. Second, carbon dioxide can bind to plasma proteins or can enter red blood cells and bind to hemoglobin. This form transports about 10 percent of the carbon dioxide. When carbon dioxide binds to hemoglobin, a molecule called **carbaminohemoglobin** is formed. Binding of carbon dioxide to hemoglobin is reversible. Therefore, when it reaches the lungs, the carbon dioxide can freely dissociate from the hemoglobin and be expelled from the body.

Third, the majority of carbon dioxide molecules (85 percent) are carried as part of the **bicarbonate buffer system**. In this system, carbon dioxide diffuses into the red blood cells. **Carbonic anhydrase (CA)** within the red blood cells quickly converts the carbon dioxide into carbonic acid (H_2CO_3). Carbonic acid is an unstable intermediate molecule that immediately dissociates into **bicarbonate ions** $\left(HCO_3^-\right)$ and hydrogen (H^+) ions. Since carbon dioxide is quickly converted into bicarbonate ions, this reaction allows for the continued uptake of carbon dioxide into the blood down its concentration gradient. It also results in the production of H^+ ions. If too much H^+ is produced, it can alter blood pH. However, hemoglobin binds to the free H^+ ions and thus limits shifts in pH. The newly synthesized bicarbonate ion is transported out of the red blood cell into the liquid component of the blood in exchange for a chloride ion (Cl^-); this is called the **chloride shift**. When the blood reaches the lungs, the bicarbonate ion is transported back into the red blood cell in exchange for the chloride ion. The H^+ ion dissociates from the hemoglobin and binds to the bicarbonate ion. This produces the carbonic acid intermediate, which is converted back into carbon dioxide through the enzymatic action of CA. The carbon dioxide produced is expelled through the lungs during exhalation.

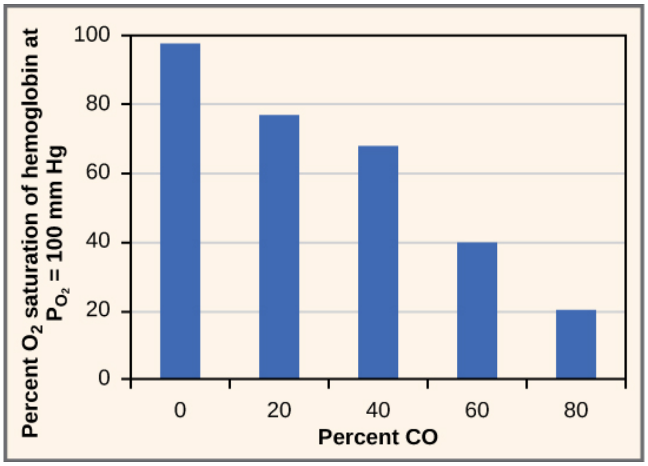
Equation:

$$CO_2 + H_2O \leftrightarrow \underset{\text{(carbonic acid)}}{H_2CO_3} \leftrightarrow \underset{\text{(bicarbonate)}}{HCO_3 + H^+}$$

The benefit of the bicarbonate buffer system is that carbon dioxide is “soaked up” into the blood with little change to the pH of the system. This is important because it takes only a small change in the overall pH of the body for severe injury or death to result. The presence of this bicarbonate buffer system also allows for people to travel and live at high altitudes: When the partial pressure of oxygen and carbon dioxide change at high altitudes, the bicarbonate buffer system adjusts to regulate carbon dioxide while maintaining the correct pH in the body.

Carbon Monoxide Poisoning

While carbon dioxide can readily associate and dissociate from hemoglobin, other molecules such as carbon monoxide (CO) cannot. Carbon monoxide has a greater affinity for hemoglobin than oxygen. Therefore, when carbon monoxide is present, it binds to hemoglobin preferentially over oxygen. As a result, oxygen cannot bind to hemoglobin, so very little oxygen is transported through the body ([link]). Carbon monoxide is a colorless, odorless gas and is therefore difficult to detect. It is produced by gas-powered vehicles and tools. Carbon monoxide can cause headaches, confusion, and nausea; long-term exposure can cause brain damage or death. Administering 100 percent (pure) oxygen is the usual treatment for carbon monoxide poisoning. Administration of pure oxygen speeds up the separation of carbon monoxide from hemoglobin.



As percent CO increases, the oxygen saturation of hemoglobin decreases.

Section Summary

Hemoglobin is a protein found in red blood cells that is comprised of two alpha and two beta subunits that surround an iron-containing heme group. Oxygen readily binds this heme group. The ability of oxygen to bind increases as more oxygen molecules are bound to heme. Disease states and altered conditions in the body can affect the binding ability of oxygen, and increase or decrease its ability to dissociate from hemoglobin.

Carbon dioxide can be transported through the blood via three methods. It is dissolved directly in the blood, bound to plasma proteins or hemoglobin, or converted into bicarbonate. The majority of carbon dioxide is transported as part of the bicarbonate system. Carbon dioxide diffuses into red blood cells. Inside, carbonic anhydrase converts carbon dioxide into carbonic acid (H_2CO_3), which is subsequently hydrolyzed into bicarbonate $\left(\mathrm{HCO}_3^-\right)$ and H^+. The H^+ ion binds to hemoglobin in red blood cells, and bicarbonate is transported out of the red blood cells in exchange for a chloride ion. This is called the chloride shift. Bicarbonate leaves the red blood cells and enters the blood plasma. In the lungs, bicarbonate is transported back into the red

blood cells in exchange for chloride. The H^+ dissociates from hemoglobin and combines with bicarbonate to form carbonic acid with the help of carbonic anhydrase, which further catalyzes the reaction to convert carbonic acid back into carbon dioxide and water. The carbon dioxide is then expelled from the lungs.

Art Connections

Exercise:

Problem:

[link] The kidneys are responsible for removing excess H+ ions from the blood. If the kidneys fail, what would happen to blood pH and to hemoglobin affinity for oxygen?

Solution:

[link] The blood pH will drop and hemoglobin affinity for oxygen will decrease.

Review Questions

Exercise:

Problem:

Which of the following will NOT facilitate the transfer of oxygen to tissues?

a. decreased body temperature
b. decreased pH of the blood
c. increased carbon dioxide
d. increased exercise

Solution:

A

Exercise:

Problem:

The majority of carbon dioxide in the blood is transported by _________.

a. binding to hemoglobin
b. dissolution in the blood
c. conversion to bicarbonate
d. binding to plasma proteins

Solution:

C

Exercise:

Problem:

The majority of oxygen in the blood is transported by _________.

a. dissolution in the blood
b. being carried as bicarbonate ions
c. binding to blood plasma
d. binding to hemoglobin

Solution:

D

Free Response

Exercise:

Problem:

What would happen if no carbonic anhydrase were present in red blood cells?

Solution:

Without carbonic anhydrase, carbon dioxide would not be hydrolyzed into carbonic acid or bicarbonate. Therefore, very little carbon dioxide (only 15 percent) would be transported in the blood away from the tissues.

Exercise:

Problem:

How does the administration of 100 percent oxygen save a patient from carbon monoxide poisoning? Why wouldn't giving carbon dioxide work?

Solution:

Carbon monoxide has a higher affinity for hemoglobin than oxygen. This means that carbon monoxide will preferentially bind to hemoglobin over oxygen. Administration of 100 percent oxygen is an effective therapy because at that concentration, oxygen will displace the carbon monoxide from the hemoglobin.

Glossary

bicarbonate buffer system
: system in the blood that absorbs carbon dioxide and regulates pH levels

bicarbonate $\left(HCO_3^-\right)$ ion
: ion created when carbonic acid dissociates into H^+ and $\left(HCO_3^-\right)$

carbaminohemoglobin
: molecule that forms when carbon dioxide binds to hemoglobin

carbonic anhydrase (CA)
: enzyme that catalyzes carbon dioxide and water into carbonic acid

chloride shift
: chloride shift exchange of chloride for bicarbonate into or out of the red blood cell

heme group
: centralized iron-containing group that is surrounded by the alpha and beta subunits of hemoglobin

hemoglobin
: molecule in red blood cells that can bind oxygen, carbon dioxide, and carbon monoxide

oxygen-carrying capacity
: amount of oxygen that can be transported in the blood

oxygen dissociation curve
: curve depicting the affinity of oxygen for hemoglobin

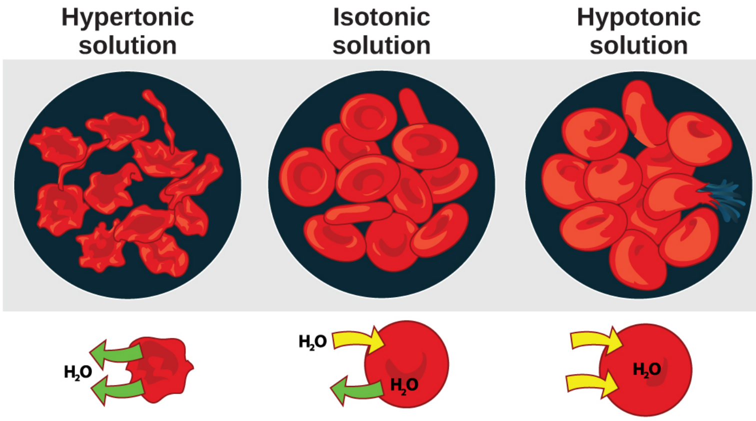
sickle cell anemia
: genetic disorder that affects the shape of red blood cells, and their ability to transport oxygen and move through capillaries

thalassemia
: rare genetic disorder that results in mutation of the alpha or beta subunits of hemoglobin, creating smaller red blood cells with less hemoglobin

Osmoregulation and Osmotic Balance
By the end of this section, you will be able to:

- Define osmosis and explain its role within molecules
- Explain why osmoregulation and osmotic balance are important body functions
- Describe active transport mechanisms
- Explain osmolarity and the way in which it is measured
- Describe osmoregulators or osmoconformers and how these tools allow animals to adapt to different environments

Osmosis is the diffusion of water across a membrane in response to **osmotic pressure** caused by an imbalance of molecules on either side of the membrane. **Osmoregulation** is the process of maintenance of salt and water balance (**osmotic balance**) across membranes within the body's fluids, which are composed of water, plus electrolytes and non-electrolytes. An **electrolyte** is a solute that dissociates into ions when dissolved in water. A **non-electrolyte**, in contrast, doesn't dissociate into ions during water dissolution. Both electrolytes and non-electrolytes contribute to the osmotic balance. The body's fluids include blood plasma, the cytosol within cells, and interstitial fluid, the fluid that exists in the spaces between cells and tissues of the body. The membranes of the body (such as the pleural, serous, and cell membranes) are **semi-permeable membranes**. Semi-permeable membranes are permeable (or permissive) to certain types of solutes and water. Solutions on two sides of a semi-permeable membrane tend to equalize in solute concentration by movement of solutes and/or water across the membrane. As seen in [link], a cell placed in water tends to swell due to gain of water from the hypotonic or "low salt" environment. A cell placed in a solution with higher salt concentration, on the other hand, tends to make the membrane shrivel up due to loss of water into the hypertonic or "high salt" environment. Isotonic cells have an equal concentration of solutes inside and outside the cell; this equalizes the osmotic pressure on either side of the cell membrane which is a semi-permeable membrane.



Cells placed in a hypertonic environment tend to shrink due to loss of water. In a hypotonic environment, cells tend to swell due to intake of water. The blood maintains an isotonic environment so that cells neither shrink nor swell. (credit: Mariana Ruiz Villareal)

The body does not exist in isolation. There is a constant input of water and electrolytes into the system. While osmoregulation is achieved across membranes within the body, excess electrolytes and wastes are transported to the kidneys and excreted, helping to maintain osmotic balance.

Need for Osmoregulation

Biological systems constantly interact and exchange water and nutrients with the environment by way of consumption of food and water and through excretion in the form of sweat, urine, and feces. Without a mechanism to regulate osmotic pressure, or when a disease damages this mechanism, there is a tendency to accumulate toxic waste and water, which can have dire consequences.

Mammalian systems have evolved to regulate not only the overall osmotic pressure across membranes, but also specific concentrations of important electrolytes in the three major fluid compartments: blood plasma,

extracellular fluid, and intracellular fluid. Since osmotic pressure is regulated by the movement of water across membranes, the volume of the fluid compartments can also change temporarily. Because blood plasma is one of the fluid components, osmotic pressures have a direct bearing on blood pressure.

Transport of Electrolytes across Cell Membranes

Electrolytes, such as sodium chloride, ionize in water, meaning that they dissociate into their component ions. In water, sodium chloride (NaCl), dissociates into the sodium ion (Na^+) and the chloride ion ($Cl^–$). The most important ions, whose concentrations are very closely regulated in body fluids, are the cations sodium (Na^+), potassium (K^+), calcium (Ca^{+2}), magnesium (Mg^{+2}), and the anions chloride (Cl^-), carbonate (CO_3^{-2}), bicarbonate (HCO_3^-), and phosphate(PO_3^-). Electrolytes are lost from the body during urination and perspiration. For this reason, athletes are encouraged to replace electrolytes and fluids during periods of increased activity and perspiration.

Osmotic pressure is influenced by the concentration of solutes in a solution. It is directly proportional to the number of solute atoms or molecules and not dependent on the size of the solute molecules. Because electrolytes dissociate into their component ions, they, in essence, add more solute particles into the solution and have a greater effect on osmotic pressure, per mass than compounds that do not dissociate in water, such as glucose.

Water can pass through membranes by passive diffusion. If electrolyte ions could passively diffuse across membranes, it would be impossible to maintain specific concentrations of ions in each fluid compartment therefore they require special mechanisms to cross the semi-permeable membranes in the body. This movement can be accomplished by facilitated diffusion and active transport. Facilitated diffusion requires protein-based channels for moving the solute. Active transport requires energy in the form of ATP conversion, carrier proteins, or pumps in order to move ions against the concentration gradient.

Concept of Osmolality and Milliequivalent

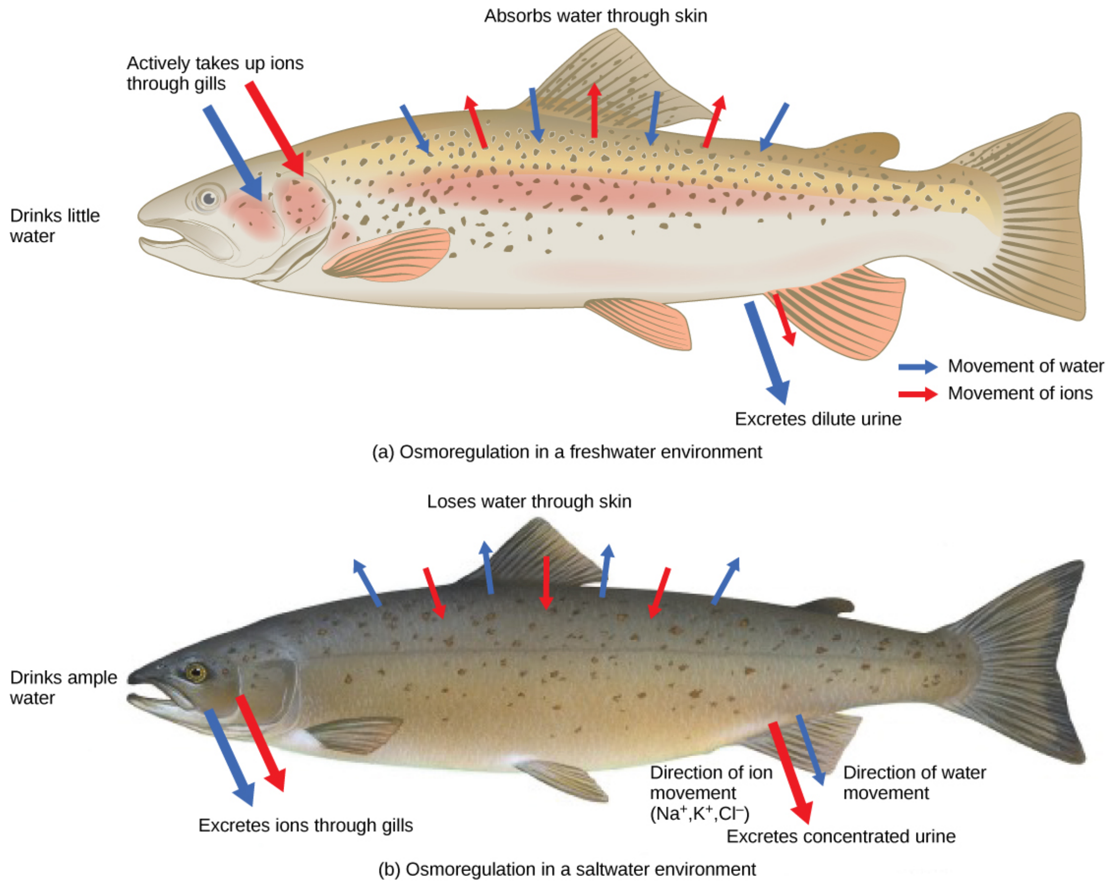
In order to calculate osmotic pressure, it is necessary to understand how solute concentrations are measured. The unit for measuring solutes is the **mole**. One mole is defined as the gram molecular weight of the solute. For example, the molecular weight of sodium chloride is 58.44. Thus, one mole of sodium chloride weighs 58.44 grams. The **molarity** of a solution is the number of moles of solute per liter of solution. The **molality** of a solution is the number of moles of solute per kilogram of solvent. If the solvent is water, one kilogram of water is equal to one liter of water. While molarity and molality are used to express the concentration of solutions, electrolyte concentrations are usually expressed in terms of milliequivalents per liter (mEq/L): the mEq/L is equal to the ion concentration (in millimoles) multiplied by the number of electrical charges on the ion. The unit of milliequivalent takes into consideration the ions present in the solution (since electrolytes form ions in aqueous solutions) and the charge on the ions.

Thus, for ions that have a charge of one, one milliequivalent is equal to one millimole. For ions that have a charge of two (like calcium), one milliequivalent is equal to 0.5 millimoles. Another unit for the expression of electrolyte concentration is the milliosmole (mOsm), which is the number of milliequivalents of solute per kilogram of solvent. Body fluids are usually maintained within the range of 280 to 300 mOsm.

Osmoregulators and Osmoconformers

Persons lost at sea without any fresh water to drink are at risk of severe dehydration because the human body cannot adapt to drinking seawater, which is hypertonic in comparison to body fluids. Organisms such as goldfish that can tolerate only a relatively narrow range of salinity are referred to as stenohaline. About 90 percent of all bony fish are restricted to either freshwater or seawater. They are incapable of osmotic regulation in the opposite environment. It is possible, however, for a few fishes like salmon to spend part of their life in fresh water and part in sea water. Organisms like the salmon and molly that can tolerate a relatively wide range of salinity are referred to as euryhaline organisms. This is possible because some fish have evolved **osmoregulatory** mechanisms to survive in all kinds of aquatic environments. When they live in fresh water, their

bodies tend to take up water because the environment is relatively hypotonic, as illustrated in [link]**a**. In such hypotonic environments, these fish do not drink much water. Instead, they pass a lot of very dilute urine, and they achieve electrolyte balance by active transport of salts through the gills. When they move to a hypertonic marine environment, these fish start drinking sea water; they excrete the excess salts through their gills and their urine, as illustrated in [link]**b**. Most marine invertebrates, on the other hand, may be isotonic with sea water (**osmoconformers**). Their body fluid concentrations conform to changes in seawater concentration. Cartilaginous fishes' salt composition of the blood is similar to bony fishes; however, the blood of sharks contains the organic compounds urea and trimethylamine oxide (TMAO). This does not mean that their electrolyte composition is similar to that of sea water. They achieve isotonicity with the sea by storing large concentrations of urea. These animals that secrete urea are called ureotelic animals. TMAO stabilizes proteins in the presence of high urea levels, preventing the disruption of peptide bonds that would occur in other animals exposed to similar levels of urea. Sharks are cartilaginous fish with a rectal gland to secrete salt and assist in osmoregulation.



Fish are osmoregulators, but must use different mechanisms to survive in (a) freshwater or (b) saltwater environments. (credit: modification of work by Duane Raver, NOAA)

Note:
Career Connection
Dialysis Technician
Dialysis is a medical process of removing wastes and excess water from the blood by diffusion and ultrafiltration. When kidney function fails, dialysis must be done to artificially rid the body of wastes. This is a vital process to keep patients alive. In some cases, the patients undergo artificial dialysis until they are eligible for a kidney transplant. In others who are not candidates for kidney transplants, dialysis is a life-long necessity.

Dialysis technicians typically work in hospitals and clinics. While some roles in this field include equipment development and maintenance, most dialysis technicians work in direct patient care. Their on-the-job duties, which typically occur under the direct supervision of a registered nurse, focus on providing dialysis treatments. This can include reviewing patient history and current condition, assessing and responding to patient needs before and during treatment, and monitoring the dialysis process. Treatment may include taking and reporting a patient's vital signs and preparing solutions and equipment to ensure accurate and sterile procedures.

Section Summary

Solute concentrations across a semi-permeable membranes influence the movement of water and solutes across the membrane. It is the number of solute molecules and not the molecular size that is important in osmosis. Osmoregulation and osmotic balance are important bodily functions, resulting in water and salt balance. Not all solutes can pass through a semi-permeable membrane. Osmosis is the movement of water across the membrane. Osmosis occurs to equalize the number of solute molecules across a semi-permeable membrane by the movement of water to the side of higher solute concentration. Facilitated diffusion utilizes protein channels to move solute molecules from areas of higher to lower concentration while active transport mechanisms are required to move solutes against concentration gradients. Osmolarity is measured in units of milliequivalents or milliosmoles, both of which take into consideration the number of solute particles and the charge on them. Fish that live in fresh water or saltwater adapt by being osmoregulators or osmoconformers.

Review Questions

Exercise:

Problem:

When a dehydrated human patient needs to be given fluids intravenously, he or she is given:

a. water, which is hypotonic with respect to body fluids
b. saline at a concentration that is isotonic with respect to body fluids
c. glucose because it is a non-electrolyte
d. blood

Solution:

B

Exercise:

Problem: The sodium ion is at the highest concentration in:

a. intracellular fluid
b. extracellular fluid
c. blood plasma
d. none of the above

Solution:

B

Exercise:

Problem: Cells in a hypertonic solution tend to:

a. shrink due to water loss
b. swell due to water gain
c. stay the same size due to water moving into and out of the cell at the same rate

d. none of the above

Solution:

A

Free Response

Exercise:

Problem:

Why is excretion important in order to achieve osmotic balance?

Solution:

Excretion allows an organism to rid itself of waste molecules that could be toxic if allowed to accumulate. It also allows the organism to keep the amount of water and dissolved solutes in balance.

Exercise:

Problem:

Why do electrolyte ions move across membranes by active transport?

Solution:

Electrolyte ions often require special mechanisms to cross the semi-permeable membranes in the body. Active transport is the movement against a concentration gradient.

Glossary

electrolyte
: solute that breaks down into ions when dissolved in water

molality
: number of moles of solute per kilogram of solvent

molarity
: number of moles of solute per liter of solution

mole
: gram equivalent of the molecular weight of a substance

non-electrolyte
: solute that does not break down into ions when dissolved in water

osmoconformer
: organism that changes its tonicity based on its environment

osmoregulation
: mechanism by which water and solute concentrations are maintained at desired levels

osmoregulator
: organism that maintains its tonicity irrespective of its environment

osmotic balance
: balance of the amount of water and salt input and output to and from a biological system without disturbing the desired osmotic pressure and solute concentration in every compartment

osmotic pressure
: pressure exerted on a membrane to equalize solute concentration on either side

semi-permeable membrane
: membrane that allows only certain solutes to pass through

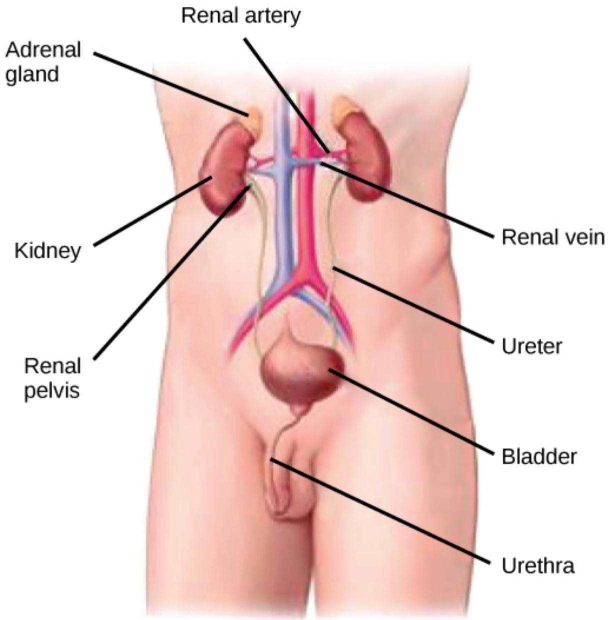
The Kidneys and Osmoregulatory Organs
By the end of this section, you will be able to:

- Explain how the kidneys serve as the main osmoregulatory organs in mammalian systems
- Describe the structure of the kidneys and the functions of the parts of the kidney
- Describe how the nephron is the functional unit of the kidney and explain how it actively filters blood and generates urine
- Detail the three steps in the formation of urine: glomerular filtration, tubular reabsorption, and tubular secretion

Although the kidneys are the major osmoregulatory organ, the skin and lungs also play a role in the process. Water and electrolytes are lost through sweat glands in the skin, which helps moisturize and cool the skin surface, while the lungs expel a small amount of water in the form of mucous secretions and via evaporation of water vapor.

Kidneys: The Main Osmoregulatory Organ

The **kidneys**, illustrated in [link], are a pair of bean-shaped structures that are located just below and posterior to the liver in the peritoneal cavity. The adrenal glands sit on top of each kidney and are also called the suprarenal glands. Kidneys filter blood and purify it. All the blood in the human body is filtered many times a day by the kidneys; these organs use up almost 25 percent of the oxygen absorbed through the lungs to perform this function. Oxygen allows the kidney cells to efficiently manufacture chemical energy in the form of ATP through aerobic respiration. The filtrate coming out of the kidneys is called **urine**.



Kidneys filter the blood, producing urine that is stored in the bladder prior to elimination through the urethra. (credit: modification of work by NCI)

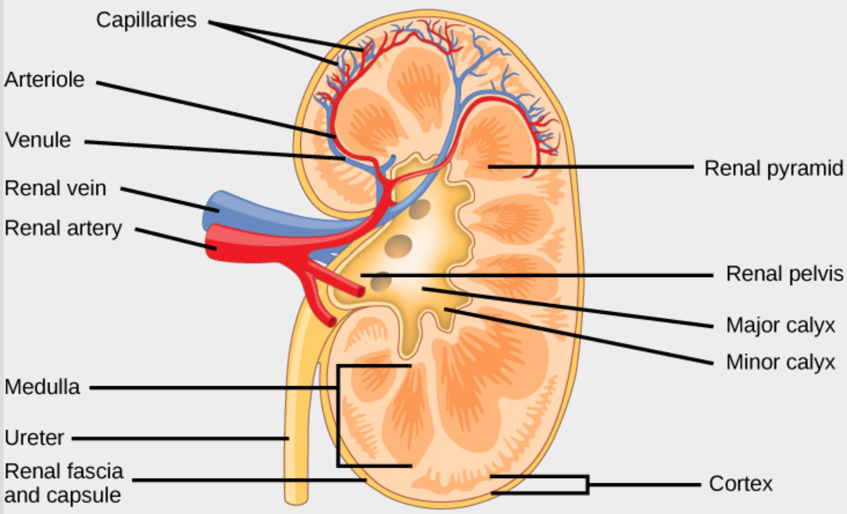
Kidney Structure

Externally, the kidneys are surrounded by three layers, illustrated in [link]. The outermost layer is a tough connective tissue layer called the **renal fascia**. The second layer is called the **perirenal fat capsule**, which helps anchor the kidneys in place. The third and innermost layer is the **renal capsule**. Internally, the kidney has three regions—an outer **cortex**, a **medulla** in the middle, and the **renal pelvis** in the region called the **hilum** of the kidney. The hilum is the concave part of the bean-shape where blood vessels and nerves enter and exit the kidney; it is also the point of exit for the ureters. The renal cortex is granular due to the presence of **nephrons**—the functional unit of the kidney. The medulla consists of multiple pyramidal tissue masses, called the **renal pyramids**. In between the pyramids are spaces called **renal columns** through which the blood vessels

pass. The tips of the pyramids, called renal papillae, point toward the renal pelvis. There are, on average, eight renal pyramids in each kidney. The renal pyramids along with the adjoining cortical region are called the **lobes of the kidney**. The renal pelvis leads to the **ureter** on the outside of the kidney. On the inside of the kidney, the renal pelvis branches out into two or three extensions called the major **calyces**, which further branch into the minor calyces. The ureters are urine-bearing tubes that exit the kidney and empty into the **urinary bladder**.

Note:

Art Connection



The internal structure of the kidney is shown. (credit: modification of work by NCI)

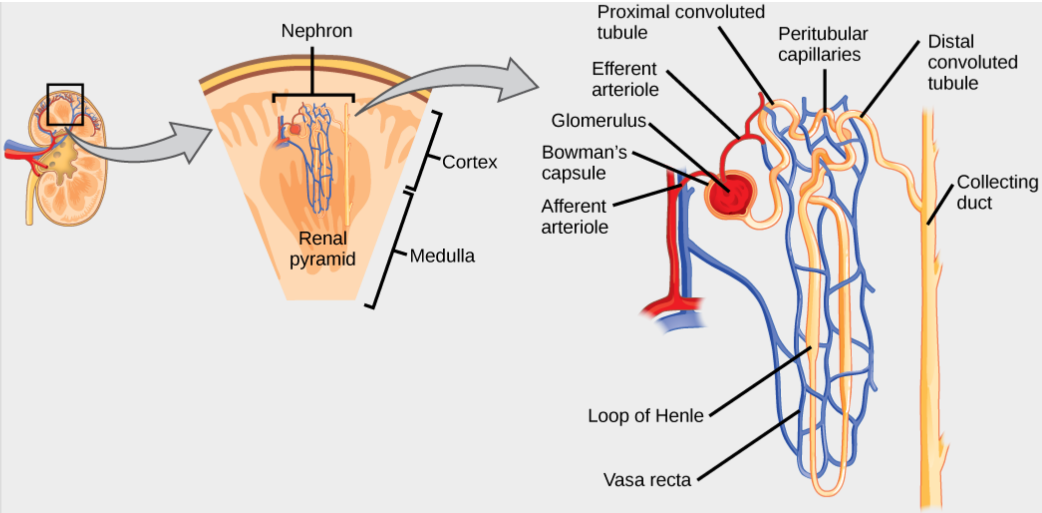
Which of the following statements about the kidney is false?

a. The renal pelvis drains into the ureter.
b. The renal pyramids are in the medulla.
c. The cortex covers the capsule.
d. Nephrons are in the renal cortex.

Because the kidney filters blood, its network of blood vessels is an important component of its structure and function. The arteries, veins, and nerves that supply the kidney enter and exit at the renal hilum. Renal blood supply starts with the branching of the aorta into the **renal arteries** (which are each named based on the region of the kidney they pass through) and ends with the exiting of the **renal veins** to join the **inferior vena cava**. The renal arteries split into several **segmental arteries** upon entering the kidneys. Each segmental artery splits further into several **interlobar arteries** and enters the renal columns, which supply the renal lobes. The interlobar arteries split at the junction of the renal cortex and medulla to form the **arcuate arteries**. The arcuate "bow shaped" arteries form arcs along the base of the medullary pyramids. **Cortical radiate arteries**, as the name suggests, radiate out from the arcuate arteries. The cortical radiate arteries branch into numerous afferent arterioles, and then enter the capillaries supplying the nephrons. Veins trace the path of the arteries and have similar names, except there are no segmental veins.

As mentioned previously, the functional unit of the kidney is the nephron, illustrated in [link]. Each kidney is made up of over one million nephrons that dot the renal cortex, giving it a granular appearance when sectioned sagittally. There are two types of nephrons—**cortical nephrons** (85 percent), which are deep in the renal cortex, and **juxtamedullary nephrons** (15 percent), which lie in the renal cortex close to the renal medulla. A nephron consists of three parts—a **renal corpuscle**, a **renal tubule**, and the associated capillary network, which originates from the cortical radiate arteries.

Note:
Art Connection



The nephron is the functional unit of the kidney. The glomerulus and convoluted tubules are located in the kidney cortex, while collecting ducts are located in the pyramids of the medulla. (credit: modification of work by NIDDK)

Which of the following statements about the nephron is false?

a. The collecting duct empties into the distal convoluted tubule.
b. The Bowman's capsule surrounds the glomerulus.
c. The loop of Henle is between the proximal and distal convoluted tubules.
d. The loop of Henle empties into the distal convoluted tubule.

Renal Corpuscle

The renal corpuscle, located in the renal cortex, is made up of a network of capillaries known as the **glomerulus** and the capsule, a cup-shaped chamber that surrounds it, called the glomerular or **Bowman's capsule**.

Renal Tubule

The renal tubule is a long and convoluted structure that emerges from the glomerulus and can be divided into three parts based on function. The first part is called the **proximal convoluted tubule (PCT)** due to its proximity to the glomerulus; it stays in the renal cortex. The second part is called the **loop of Henle**, or nephritic loop, because it forms a loop (with **descending** and **ascending limbs**) that goes through the renal medulla. The third part of the renal tubule is called the **distal convoluted tubule (DCT)** and this part is also restricted to the renal cortex. The DCT, which is the last part of the nephron, connects and empties its contents into collecting ducts that line the medullary pyramids. The collecting ducts amass contents from multiple nephrons and fuse together as they enter the papillae of the renal medulla.

Capillary Network within the Nephron

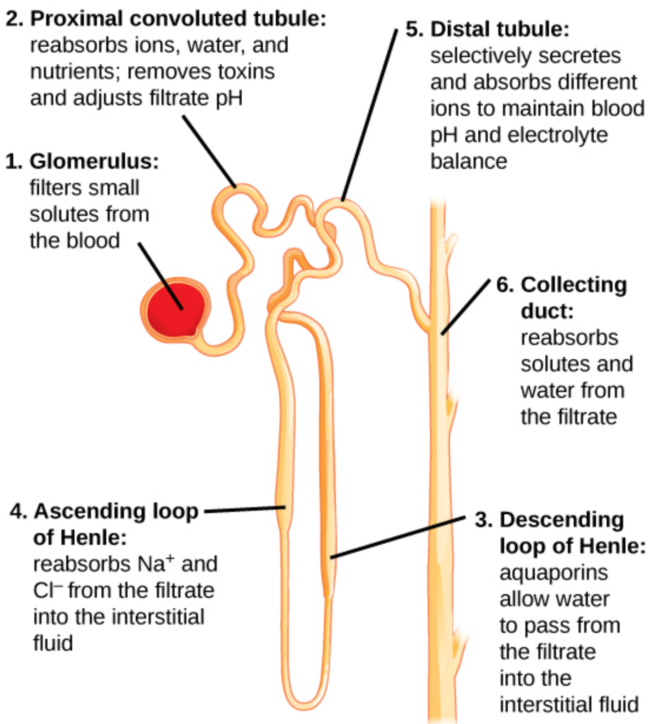
The capillary network that originates from the renal arteries supplies the nephron with blood that needs to be filtered. The branch that enters the glomerulus is called the **afferent arteriole**. The branch that exits the glomerulus is called the **efferent arteriole**. Within the glomerulus, the network of capillaries is called the glomerular capillary bed. Once the efferent arteriole exits the glomerulus, it forms the **peritubular capillary network**, which surrounds and interacts with parts of the renal tubule. In cortical nephrons, the peritubular capillary network surrounds the PCT and DCT. In juxtamedullary nephrons, the peritubular capillary network forms a network around the loop of Henle and is called the **vasa recta**.

Note:
Link to Learning

Go to this website to see another coronal section of the kidney and to explore an animation of the workings of nephrons.

Kidney Function and Physiology

Kidneys filter blood in a three-step process. First, the nephrons filter blood that runs through the capillary network in the glomerulus. Almost all solutes, except for proteins, are filtered out into the glomerulus by a process called **glomerular filtration**. Second, the filtrate is collected in the renal tubules. Most of the solutes get reabsorbed in the PCT by a process called **tubular reabsorption**. In the loop of Henle, the filtrate continues to exchange solutes and water with the renal medulla and the peritubular capillary network. Water is also reabsorbed during this step. Then, additional solutes and wastes are secreted into the kidney tubules during **tubular secretion**, which is, in essence, the opposite process to tubular reabsorption. The collecting ducts collect filtrate coming from the nephrons and fuse in the medullary papillae. From here, the papillae deliver the filtrate, now called urine, into the minor calyces that eventually connect to the ureters through the renal pelvis. This entire process is illustrated in [link].



Each part of the nephron performs a different function in filtering waste and maintaining homeostatic balance. (1) The glomerulus forces small solutes out of the blood by pressure. (2) The proximal convoluted tubule reabsorbs ions, water, and nutrients from the filtrate into the interstitial fluid, and actively transports toxins and drugs from the interstitial fluid into the filtrate. The proximal convoluted tubule also adjusts blood pH by selectively secreting ammonia (NH_3) into the filtrate, where it reacts with H^+ to form NH_4^+. The more acidic the filtrate, the more ammonia is secreted. (3) The descending loop of Henle is lined

with cells containing aquaporins that allow water to pass from the filtrate into the interstitial fluid. (4) In the thin part of the ascending loop of Henle, Na^+ and Cl^- ions diffuse into the interstitial fluid. In the thick part, these same ions are actively transported into the interstitial fluid. Because salt but not water is lost, the filtrate becomes more dilute as it travels up the limb. (5) In the distal convoluted tubule, K^+ and H^+ ions are selectively secreted into the filtrate, while Na^+, Cl^-, and HCO_3^- ions are reabsorbed to maintain pH and electrolyte balance in the blood. (6) The collecting duct reabsorbs solutes and water from the filtrate, forming dilute urine. (credit: modification of work by NIDDK)

Glomerular Filtration

Glomerular filtration filters out most of the solutes due to high blood pressure and specialized membranes in the afferent arteriole. The blood pressure in the glomerulus is maintained independent of factors that affect systemic blood pressure. The “leaky” connections between the endothelial cells of the glomerular capillary network allow solutes to pass through easily. All solutes in the glomerular capillaries, except for macromolecules like proteins, pass through by passive diffusion. There is no energy requirement at this stage of the filtration process. **Glomerular filtration**

rate (GFR) is the volume of glomerular filtrate formed per minute by the kidneys. GFR is regulated by multiple mechanisms and is an important indicator of kidney function.

Note:
Link to Learning

To learn more about the vascular system of kidneys, click through this review and the steps of blood flow.

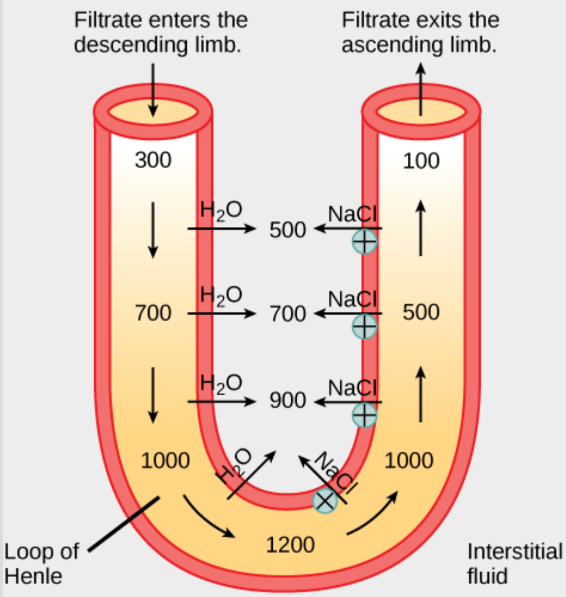
Tubular Reabsorption and Secretion

Tubular reabsorption occurs in the PCT part of the renal tubule. Almost all nutrients are reabsorbed, and this occurs either by passive or active transport. Reabsorption of water and some key electrolytes are regulated and can be influenced by hormones. Sodium (Na^+) is the most abundant ion and most of it is reabsorbed by active transport and then transported to the peritubular capillaries. Because Na^+ is actively transported out of the tubule, water follows it to even out the osmotic pressure. Water is also independently reabsorbed into the peritubular capillaries due to the presence of aquaporins, or water channels, in the PCT. This occurs due to the low blood pressure and high osmotic pressure in the peritubular capillaries. However, every solute has a **transport maximum** and the excess is not reabsorbed.

In the loop of Henle, the permeability of the membrane changes. The descending limb is permeable to water, not solutes; the opposite is true for the ascending limb. Additionally, the loop of Henle invades the renal

medulla, which is naturally high in salt concentration and tends to absorb water from the renal tubule and concentrate the filtrate. The osmotic gradient increases as it moves deeper into the medulla. Because two sides of the loop of Henle perform opposing functions, as illustrated in [link], it acts as a **countercurrent multiplier**. The vasa recta around it acts as the **countercurrent exchanger**.

Note:
Art Connection



The loop of Henle acts as a countercurrent multiplier that uses energy to create concentration gradients. The descending limb is water permeable. Water flows from the filtrate to the interstitial fluid, so osmolality inside the limb increases as it descends into the renal medulla. At the bottom, the osmolality is higher inside the loop than in

the interstitial fluid. Thus, as filtrate enters the ascending limb, Na^+ and Cl^- ions exit through ion channels present in the cell membrane. Further up, Na^+ is actively transported out of the filtrate and Cl^- follows. Osmolarity is given in units of milliosmoles per liter (mOsm/L).

Loop diuretics are drugs sometimes used to treat hypertension. These drugs inhibit the reabsorption of Na^+ and Cl^- ions by the ascending limb of the loop of Henle. A side effect is that they increase urination. Why do you think this is the case?

By the time the filtrate reaches the DCT, most of the urine and solutes have been reabsorbed. If the body requires additional water, all of it can be reabsorbed at this point. Further reabsorption is controlled by hormones, which will be discussed in a later section. Excretion of wastes occurs due to lack of reabsorption combined with tubular secretion. Undesirable products like metabolic wastes, urea, uric acid, and certain drugs, are excreted by tubular secretion. Most of the tubular secretion happens in the DCT, but some occurs in the early part of the collecting duct. Kidneys also maintain an acid-base balance by secreting excess H^+ ions.

Although parts of the renal tubules are named proximal and distal, in a cross-section of the kidney, the tubules are placed close together and in contact with each other and the glomerulus. This allows for exchange of chemical messengers between the different cell types. For example, the DCT ascending limb of the loop of Henle has masses of cells called **macula densa**, which are in contact with cells of the afferent arterioles called **juxtaglomerular cells**. Together, the macula densa and juxtaglomerular

cells form the juxtaglomerular complex (JGC). The JGC is an endocrine structure that secretes the enzyme renin and the hormone erythropoietin. When hormones trigger the macula densa cells in the DCT due to variations in blood volume, blood pressure, or electrolyte balance, these cells can immediately communicate the problem to the capillaries in the afferent and efferent arterioles, which can constrict or relax to change the glomerular filtration rate of the kidneys.

Note:
Career Connection
Nephrologist
A nephrologist studies and deals with diseases of the kidneys—both those that cause kidney failure (such as diabetes) and the conditions that are produced by kidney disease (such as hypertension). Blood pressure, blood volume, and changes in electrolyte balance come under the purview of a nephrologist.
Nephrologists usually work with other physicians who refer patients to them or consult with them about specific diagnoses and treatment plans. Patients are usually referred to a nephrologist for symptoms such as blood or protein in the urine, very high blood pressure, kidney stones, or renal failure.
Nephrology is a subspecialty of internal medicine. To become a nephrologist, medical school is followed by additional training to become certified in internal medicine. An additional two or more years is spent specifically studying kidney disorders and their accompanying effects on the body.

Section Summary

The kidneys are the main osmoregulatory organs in mammalian systems; they function to filter blood and maintain the osmolarity of body fluids at 300 mOsm. They are surrounded by three layers and are made up internally of three distinct regions—the cortex, medulla, and pelvis.

The blood vessels that transport blood into and out of the kidneys arise from and merge with the aorta and inferior vena cava, respectively. The renal arteries branch out from the aorta and enter the kidney where they further divide into segmental, interlobar, arcuate, and cortical radiate arteries.

The nephron is the functional unit of the kidney, which actively filters blood and generates urine. The nephron is made up of the renal corpuscle and renal tubule. Cortical nephrons are found in the renal cortex, while juxtamedullary nephrons are found in the renal cortex close to the renal medulla. The nephron filters and exchanges water and solutes with two sets of blood vessels and the tissue fluid in the kidneys.

There are three steps in the formation of urine: glomerular filtration, which occurs in the glomerulus; tubular reabsorption, which occurs in the renal tubules; and tubular secretion, which also occurs in the renal tubules.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the kidney is false?

a. The renal pelvis drains into the ureter.
b. The renal pyramids are in the medulla.
c. The cortex covers the capsule.
d. Nephrons are in the renal cortex.

Solution:

[link] C

Exercise:

Problem:

[link] Which of the following statements about the nephron is false?

a. The collecting duct empties into the distal convoluted tubule.
b. The Bowman’s capsule surrounds the glomerulus.
c. The loop of Henle is between the proximal and distal convoluted tubules.
d. The loop of Henle empties into the distal convoluted tubule.

Solution:

[link] A

Exercise:

Problem:

[link] Loop diuretics are drugs sometimes used to treat hypertension. These drugs inhibit the reabsorption of Na^+ and Cl^- ions by the ascending limb of the loop of Henle. A side effect is that they increase urination. Why do you think this is the case?

Solution:

[link] Loop diuretics decrease the excretion of salt into the renal medulla, thereby reducing its osmolality. As a result, less water is excreted into the medulla by the descending limb, and more water is excreted as urine.

Review Questions

Exercise:

Problem: The macula densa is/are:

a. present in the renal medulla.
b. dense tissue present in the outer layer of the kidney.
c. cells present in the DCT and collecting tubules.
d. present in blood capillaries.

Solution:

C

Exercise:

Problem: The osmolarity of body fluids is maintained at ________.

a. 100 mOsm
b. 300 mOsm
c. 1000 mOsm
d. it is not constantly maintained

Solution:

B

Exercise:

Problem:

The gland located at the top of the kidney is the ________ gland.

a. adrenal
b. pituitary
c. thyroid
d. thymus

Solution:

A

Free Response

Exercise:

Problem:

Why are the loop of Henle and vasa recta important for the formation of concentrated urine?

Solution:

The loop of Henle is part of the renal tubule that loops into the renal medulla. In the loop of Henle, the filtrate exchanges solutes and water with the renal medulla and the vasa recta (the peritubular capillary network). The vasa recta acts as the countercurrent exchanger. The kidneys maintain the osmolality of the rest of the body at a constant 300 mOsm by concentrating the filtrate as it passes through the loop of Henle.

Exercise:

Problem: Describe the structure of the kidney.

Solution:

Externally, the kidneys are surrounded by three layers. The outermost layer is a tough connective tissue layer called the renal fascia. The second layer is called the perirenal fat capsule, which helps anchor the kidneys in place. The third and innermost layer is the renal capsule. Internally, the kidney has three regions—an outer cortex, a medulla in the middle, and the renal pelvis in the region called the hilum of the kidney, which is the concave part of the "bean" shape.

Glossary

afferent arteriole
: arteriole that branches from the cortical radiate artery and enters the glomerulus

arcuate artery

artery that branches from the interlobar artery and arches over the base of the renal pyramids

ascending limb
: part of the loop of Henle that ascends from the renal medulla to the renal cortex

Bowman's capsule
: structure that encloses the glomerulus

calyx
: structure that connects the renal pelvis to the renal medulla

cortex (animal)
: outer layer of an organ like the kidney or adrenal gland

cortical nephron
: nephron that lies in the renal cortex

cortical radiate artery
: artery that radiates from the arcuate arteries into the renal cortex

countercurrent exchanger
: peritubular capillary network that allows exchange of solutes and water from the renal tubules

countercurrent multiplier
: osmotic gradient in the renal medulla that is responsible for concentration of urine

descending limb
: part of the loop of Henle that descends from the renal cortex into the renal medulla

distal convoluted tubule (DCT)
: part of the renal tubule that is the most distant from the glomerulus

efferent arteriole
: arteriole that exits from the glomerulus

glomerular filtration
filtration of blood in the glomerular capillary network into the glomerulus

glomerular filtration rate (GFR)
amount of filtrate formed by the glomerulus per minute

glomerulus (renal)
part of the renal corpuscle that contains the capillary network

hilum
region in the renal pelvis where blood vessels, nerves, and ureters bunch before entering or exiting the kidney

inferior vena cava
one of the main veins in the human body

interlobar artery
artery that branches from the segmental artery and travels in between the renal lobes

juxtaglomerular cell
cell in the afferent and efferent arterioles that responds to stimuli from the macula densa

juxtamedullary nephron
nephron that lies in the cortex but close to the renal medulla

kidney
organ that performs excretory and osmoregulatory functions

lobes of the kidney
renal pyramid along with the adjoining cortical region

loop of Henle
part of the renal tubule that loops into the renal medulla

macula densa

group of cells that senses changes in sodium ion concentration; present in parts of the renal tubule and collecting ducts

medulla
: middle layer of an organ like the kidney or adrenal gland

nephron
: functional unit of the kidney

perirenal fat capsule
: fat layer that suspends the kidneys

peritubular capillary network
: capillary network that surrounds the renal tubule after the efferent artery exits the glomerulus

proximal convoluted tubule (PCT)
: part of the renal tubule that lies close to the glomerulus

renal artery
: branch of the artery that enters the kidney

renal capsule
: layer that encapsulates the kidneys

renal column
: area of the kidney through which the interlobar arteries travel in the process of supplying blood to the renal lobes

renal corpuscle
: glomerulus and the Bowman's capsule together

renal fascia
: connective tissue that supports the kidneys

renal pelvis
: region in the kidney where the calyces join the ureters

renal pyramid

conical structure in the renal medulla

renal tubule
: tubule of the nephron that arises from the glomerulus

renal vein
: branch of a vein that exits the kidney and joins the inferior vena cava

segmental artery
: artery that branches from the renal artery

transport maximum
: maximum amount of solute that can be transported out of the renal tubules during reabsorption

tubular reabsorption
: reclamation of water and solutes that got filtered out in the glomerulus

tubular secretion
: process of secretion of wastes that do not get reabsorbed

ureter
: urine-bearing tube coming out of the kidney; carries urine to the bladder

urinary bladder
: structure that the ureters empty the urine into; stores urine

urine
: filtrate produced by kidneys that gets excreted out of the body

vasa recta
: peritubular network that surrounds the loop of Henle of the juxtamedullary nephrons

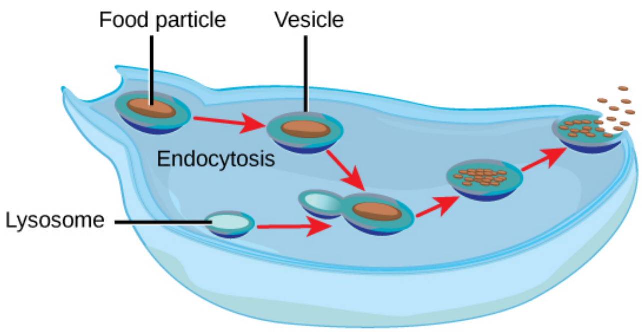
Excretion Systems
By the end of this section, you will be able to:

- Explain how vacuoles, present in microorganisms, work to excrete waste
- Describe the way in which flame cells and nephridia in worms perform excretory functions and maintain osmotic balance
- Explain how insects use Malpighian tubules to excrete wastes and maintain osmotic balance

Microorganisms and invertebrate animals use more primitive and simple mechanisms to get rid of their metabolic wastes than the mammalian system of kidney and urinary function. Three excretory systems evolved in organisms before complex kidneys: vacuoles, flame cells, and Malpighian tubules.

Contractile Vacuoles in Microorganisms

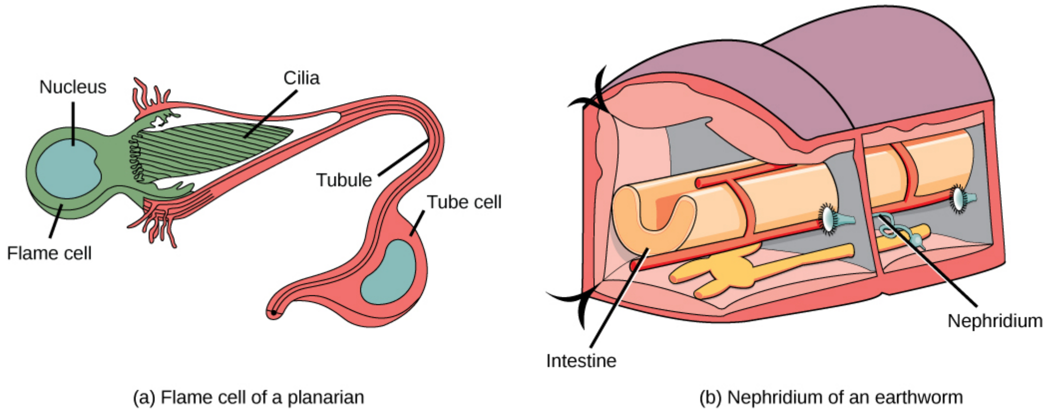
The most fundamental feature of life is the presence of a cell. In other words, a cell is the simplest functional unit of a life. Bacteria are unicellular, prokaryotic organisms that have some of the least complex life processes in place; however, prokaryotes such as bacteria do not contain membrane-bound vacuoles. The cells of microorganisms like bacteria, protozoa, and fungi are bound by cell membranes and use them to interact with the environment. Some cells, including some leucocytes in humans, are able to engulf food by endocytosis—the formation of vesicles by involution of the cell membrane within the cells. The same vesicles are able to interact and exchange metabolites with the intracellular environment. In some unicellular eukaryotic organisms such as the amoeba, shown in [link], cellular wastes and excess water are excreted by exocytosis, when the contractile vacuoles merge with the cell membrane and expel wastes into the environment. Contractile vacuoles (CV) should not be confused with vacuoles, which store food or water.



Some unicellular organisms, such as the amoeba, ingest food by endocytosis. The food vesicle fuses with a lysosome, which digests the food. Waste is excreted by exocytosis.

Flame Cells of Planaria and Nephridia of Worms

As multi-cellular systems evolved to have organ systems that divided the metabolic needs of the body, individual organs evolved to perform the excretory function. Planaria are flatworms that live in fresh water. Their excretory system consists of two tubules connected to a highly branched duct system. The cells in the tubules are called **flame cells** (or **protonephridia**) because they have a cluster of cilia that looks like a flickering flame when viewed under the microscope, as illustrated in [link]**a**. The cilia propel waste matter down the tubules and out of the body through excretory pores that open on the body surface; cilia also draw water from the interstitial fluid, allowing for filtration. Any valuable metabolites are recovered by reabsorption. Flame cells are found in flatworms, including parasitic tapeworms and free-living planaria. They also maintain the organism’s osmotic balance.



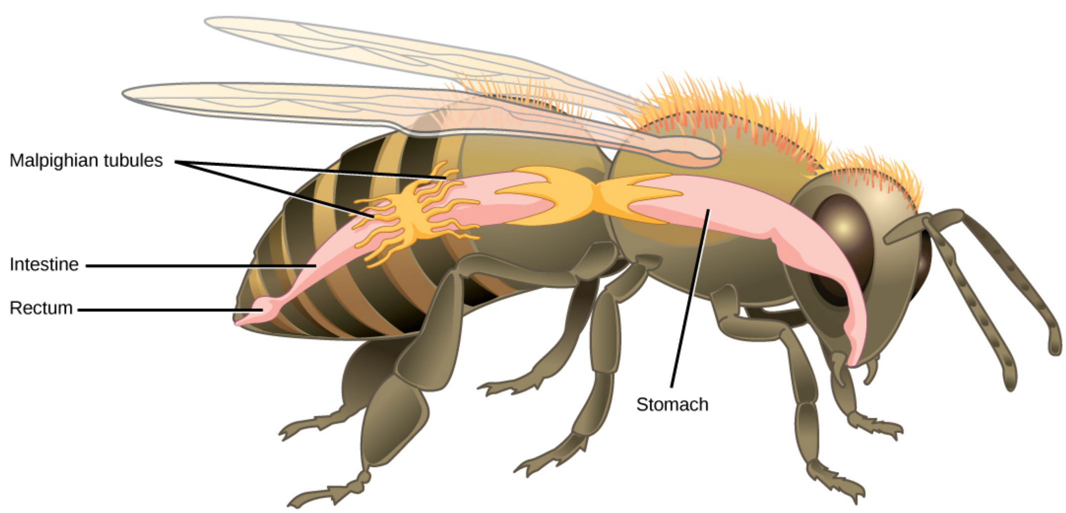
In the excretory system of the (a) planaria, cilia of flame cells propel waste through a tubule formed by a tube cell. Tubules are connected into branched structures that lead to pores located all along the sides of the body. The filtrate is secreted through these pores. In (b) annelids such as earthworms, nephridia filter fluid from the coelom, or body cavity. Beating cilia at the opening of the nephridium draw water from the coelom into a tubule. As the filtrate passes down the tubules, nutrients and other solutes are reabsorbed by capillaries. Filtered fluid containing nitrogenous and other wastes is stored in a bladder and then secreted through a pore in the side of the body.

Earthworms (annelids) have slightly more evolved excretory structures called **nephridia**, illustrated in [link]**b**. A pair of nephridia is present on each segment of the earthworm. They are similar to flame cells in that they have a tubule with cilia. Excretion occurs through a pore called the **nephridiopore**. They are more evolved than the flame cells in that they have a system for tubular reabsorption by a capillary network before excretion.

Malpighian Tubules of Insects

Malpighian tubules are found lining the gut of some species of arthropods, such as the bee illustrated in [link]. They are usually found in pairs and the

number of tubules varies with the species of insect. Malpighian tubules are convoluted, which increases their surface area, and they are lined with **microvilli** for reabsorption and maintenance of osmotic balance. Malpighian tubules work cooperatively with specialized glands in the wall of the rectum. Body fluids are not filtered as in the case of nephridia; urine is produced by tubular secretion mechanisms by the cells lining the Malpighian tubules that are bathed in hemolymph (a mixture of blood and interstitial fluid that is found in insects and other arthropods as well as most mollusks). Metabolic wastes like uric acid freely diffuse into the tubules. There are exchange pumps lining the tubules, which actively transport H^+ ions into the cell and K^+ or Na^+ ions out; water passively follows to form urine. The secretion of ions alters the osmotic pressure which draws water, electrolytes, and nitrogenous waste (uric acid) into the tubules. Water and electrolytes are reabsorbed when these organisms are faced with low-water environments, and uric acid is excreted as a thick paste or powder. Not dissolving wastes in water helps these organisms to conserve water; this is especially important for life in dry environments.



Malpighian tubules of insects and other terrestrial arthropods remove nitrogenous wastes and other solutes from the hemolymph. Na^+ and/or K^+ ions are actively transported into the lumen of the tubules. Water then enters the tubules via osmosis, forming urine. The urine passes through the intestine,

and into the rectum. There, nutrients diffuse back into the hemolymph. Na^+ and/or K^+ ions are pumped into the hemolymph, and water follows. The concentrated waste is then excreted.

Note:
Link to Learning

See a dissected cockroach, including a close-up look at its Malpighian tubules, in this video.

Section Summary

Many systems have evolved for excreting wastes that are simpler than the kidney and urinary systems of vertebrate animals. The simplest system is that of contractile vacuoles present in microorganisms. Flame cells and nephridia in worms perform excretory functions and maintain osmotic balance. Some insects have evolved Malpighian tubules to excrete wastes and maintain osmotic balance.

Review Questions

Exercise:

Problem: Active transport of K^+ in Malpighian tubules ensures that:

a. water follows K^+ to make urine
b. osmotic balance is maintained between waste matter and bodily fluids
c. both a and b
d. neither a nor b

Solution:

C

Exercise:

Problem: Contractile vacuoles in microorganisms:

a. exclusively perform an excretory function
b. can perform many functions, one of which is excretion of metabolic wastes
c. originate from the cell membrane
d. both b and c

Solution:

D

Exercise:

Problem:

Flame cells are primitive excretory organs found in ________.

a. arthropods
b. annelids
c. mammals
d. flatworms

Solution:

D

Free Response

Exercise:

Problem:

Why might specialized organs have evolved for excretion of wastes?

Solution:

The removal of wastes, which could otherwise be toxic to an organism, is extremely important for survival. Having organs that specialize in this process and that operate separately from other organs provides a measure of safety for the organism.

Exercise:

Problem:

Explain two different excretory systems other than the kidneys.

Solution:

(1) Microorganisms engulf food by endocytosis—the formation of vacuoles by involution of the cell membrane within the cells. The same vacuoles interact and exchange metabolites with the intracellular environment. Cellular wastes are excreted by exocytosis when the vacuoles merge with the cell membrane and excrete wastes into the environment. (2) Flatworms have an excretory system that consists of two tubules. The cells in the tubules are called flame cells; they have a cluster of cilia that propel waste matter down the tubules and out of the body. (3) Annelids have nephridia which have a tubule with cilia. Excretion occurs through a pore called the nephridiopore. Annelids have a system for tubular reabsorption by a capillary network before excretion. (4) Malpighian tubules are found in some species of arthropods. They are usually found in pairs, and the number of tubules varies with the species of insect. Malpighian tubules are convoluted,

which increases their surface area, and they are lined with microvilli for reabsorption and maintenance of osmotic balance. Metabolic wastes like uric acid freely diffuse into the tubules. Potassium ion pumps line the tubules, which actively transport out K^+ ions, and water follows to form urine. Water and electrolytes are reabsorbed when these organisms are faced with low-water environments, and uric acid is excreted as a thick paste or powder. By not dissolving wastes in water, these organisms conserve water.

Glossary

flame cell
: (also, protonephridia) excretory cell found in flatworms

Malpighian tubule
: excretory tubules found in arthropods

microvilli
: cellular processes that increase the surface area of cells

nephridia
: excretory structures found in annelids

nephridiopore
: pore found at the end of nephridia

Nitrogenous Wastes
By the end of this section, you will be able to:

- Compare and contrast the way in which aquatic animals and terrestrial animals can eliminate toxic ammonia from their systems
- Compare the major byproduct of ammonia metabolism in vertebrate animals to that of birds, insects, and reptiles

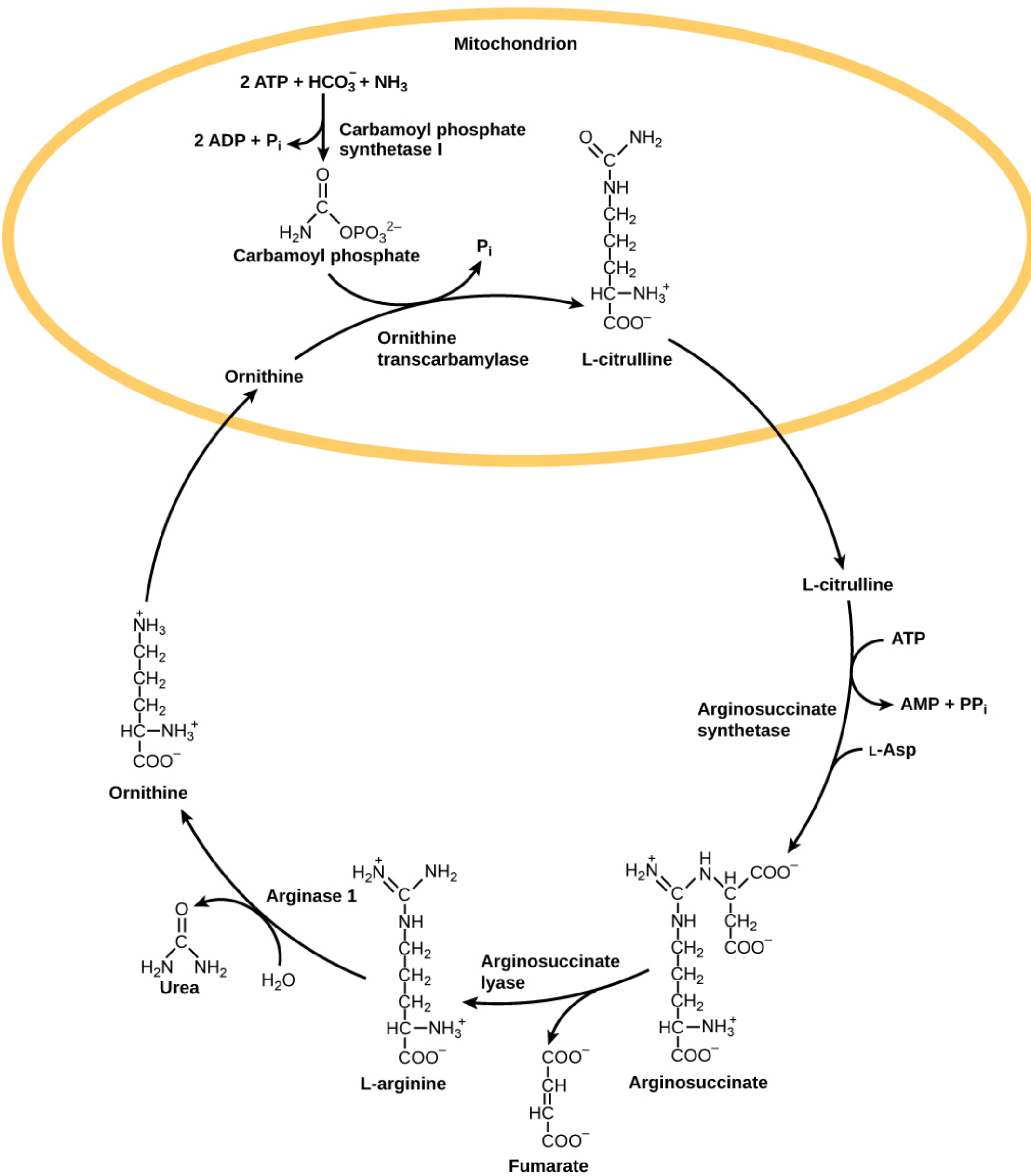
Of the four major macromolecules in biological systems, both proteins and nucleic acids contain nitrogen. During the catabolism, or breakdown, of nitrogen-containing macromolecules, carbon, hydrogen, and oxygen are extracted and stored in the form of carbohydrates and fats. Excess nitrogen is excreted from the body. Nitrogenous wastes tend to form toxic **ammonia**, which raises the pH of body fluids. The formation of ammonia itself requires energy in the form of ATP and large quantities of water to dilute it out of a biological system. Animals that live in aquatic environments tend to release ammonia into the water. Animals that excrete ammonia are said to be **ammonotelic**. Terrestrial organisms have evolved other mechanisms to excrete nitrogenous wastes. The animals must detoxify ammonia by converting it into a relatively nontoxic form such as urea or uric acid. Mammals, including humans, produce urea, whereas reptiles and many terrestrial invertebrates produce uric acid. Animals that secrete urea as the primary nitrogenous waste material are called **ureotelic** animals.

Nitrogenous Waste in Terrestrial Animals: The Urea Cycle

The **urea cycle** is the primary mechanism by which mammals convert ammonia to urea. Urea is made in the liver and excreted in urine. The overall chemical reaction by which ammonia is converted to urea is $2\ NH_3$ (ammonia) + CO_2 + 3 ATP + H_2O → $H_2N\text{-}CO\text{-}NH_2$ (urea) + 2 ADP + 4 P_i + AMP.

The urea cycle utilizes five intermediate steps, catalyzed by five different enzymes, to convert ammonia to urea, as shown in [link]. The amino acid L-ornithine gets converted into different intermediates before being regenerated at the end of the urea cycle. Hence, the urea cycle is also referred to as the ornithine cycle. The enzyme ornithine transcarbamylase

catalyzes a key step in the urea cycle and its deficiency can lead to accumulation of toxic levels of ammonia in the body. The first two reactions occur in the mitochondria and the last three reactions occur in the cytosol. Urea concentration in the blood, called **blood urea nitrogen** or BUN, is used as an indicator of kidney function.

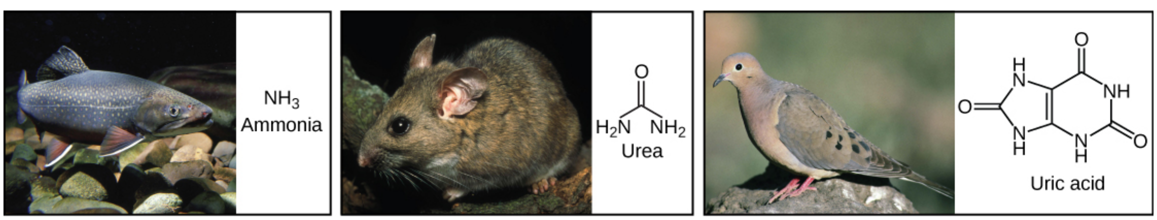


The urea cycle converts ammonia to urea.

Note:
Evolution Connection
Excretion of Nitrogenous Waste
The theory of evolution proposes that life started in an aquatic environment. It is not surprising to see that biochemical pathways like the urea cycle evolved to adapt to a changing environment when terrestrial life forms evolved. Arid conditions probably led to the evolution of the uric acid pathway as a means of conserving water.

Nitrogenous Waste in Birds and Reptiles: Uric Acid

Birds, reptiles, and most terrestrial arthropods convert toxic ammonia to **uric acid** or the closely related compound guanine (guano) instead of urea. Mammals also form some uric acid during breakdown of nucleic acids. Uric acid is a compound similar to purines found in nucleic acids. It is water insoluble and tends to form a white paste or powder; it is excreted by birds, insects, and reptiles. Conversion of ammonia to uric acid requires more energy and is much more complex than conversion of ammonia to urea [link].



(a) Many invertebrates and aquatic species excrete ammonia.
(b) Mammals, many adult amphibians, and some marine species excrete urea.
(c) Insects, land snails, birds, and many reptiles excrete uric acid.

Nitrogenous waste is excreted in different forms by different species. These include (a) ammonia, (b) urea, and (c) uric acid. (credit a: modification of work by Eric Engbretson, USFWS;

credit b: modification of work by B. "Moose" Peterson, USFWS; credit c: modification of work by Dave Menke, USFWS)

Note:
Everyday Connection

Gout

Mammals use uric acid crystals as an **antioxidant** in their cells. However, too much uric acid tends to form kidney stones and may also cause a painful condition called gout, where uric acid crystals accumulate in the joints, as illustrated in [link]. Food choices that reduce the amount of nitrogenous bases in the diet help reduce the risk of gout. For example, tea, coffee, and chocolate have purine-like compounds, called xanthines, and should be avoided by people with gout and kidney stones.

Gout causes the inflammation visible in this person’s left big toe joint. (credit: "Gonzosft"/Wikimedia Commons)

Section Summary

Ammonia is the waste produced by metabolism of nitrogen-containing compounds like proteins and nucleic acids. While aquatic animals can easily excrete ammonia into their watery surroundings, terrestrial animals have evolved special mechanisms to eliminate the toxic ammonia from their systems. Urea is the major byproduct of ammonia metabolism in vertebrate animals. Uric acid is the major byproduct of ammonia metabolism in birds, terrestrial arthropods, and reptiles.

Review Questions

Exercise:

Problem: BUN is _________.

a. blood urea nitrogen
b. blood uric acid nitrogen
c. an indicator of blood volume
d. an indicator of blood pressure

Solution:

A

Exercise:

Problem:

Human beings accumulate _________ before excreting nitrogenous waste.

a. nitrogen
b. ammonia
c. urea
d. uric acid

Solution:

C

Free Response

Exercise:

Problem:

In terms of evolution, why might the urea cycle have evolved in organisms?

Solution:

It is believed that the urea cycle evolved to adapt to a changing environment when terrestrial life forms evolved. Arid conditions probably led to the evolution of the uric acid pathway as a means of conserving water.

Exercise:

Problem: Compare and contrast the formation of urea and uric acid.

Solution:

The urea cycle is the primary mechanism by which mammals convert ammonia to urea. Urea is made in the liver and excreted in urine. The urea cycle utilizes five intermediate steps, catalyzed by five different enzymes, to convert ammonia to urea. Birds, reptiles, and insects, on the other hand, convert toxic ammonia to uric acid instead of urea. Conversion of ammonia to uric acid requires more energy and is much more complex than conversion of ammonia to urea.

Glossary

ammonia

compound made of one nitrogen atom and three hydrogen atoms

ammonotelic
: describes an animal that excretes ammonia as the primary waste material

antioxidant
: agent that prevents cell destruction by reactive oxygen species

blood urea nitrogen (BUN)
: estimate of urea in the blood and an indicator of kidney function

urea cycle
: pathway by which ammonia is converted to urea

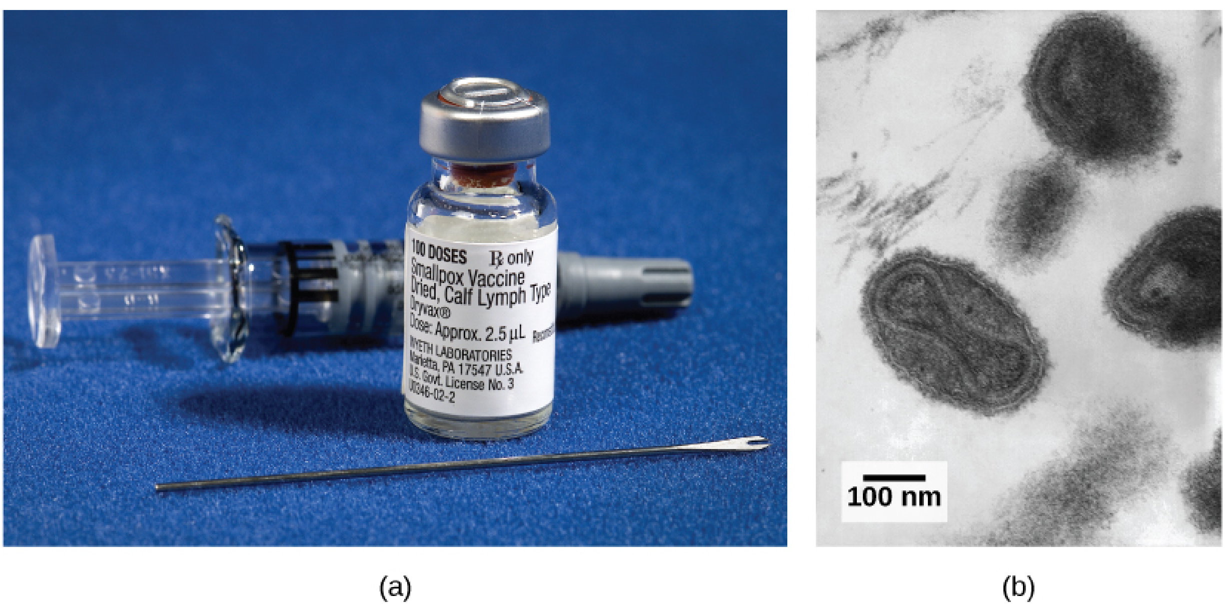
ureotelic
: describes animals that secrete urea as the primary nitrogenous waste material

uric acid
: byproduct of ammonia metabolism in birds, insects, and reptiles

Introduction
class="introduction"

(a) This smallpox (variola) vaccine is derived from calves exposed to cowpox virus. Vaccines provoke a reaction in the immune system that prepares it for a subsequent infection by smallpox. (b) Viewed under a transmission electron microscope, you can see the variola’s dumbbell-shaped structure that contains the viral DNA. (credit a: modificatio

n of work by James Gathany, CDC; credit b: modification of work by Dr. Fred Murphy; Sylvia Whitfield, CDC; scale-bar data from Matt Russell)



Organisms have a wide array of adaptations for preventing attacks of parasites and diseases. The vertebrate defense systems, including those of humans, are complex and multilayered, with defenses unique to vertebrates. These unique vertebrate defenses interact with other defense systems inherited from ancestral lineages, and include complex and specific

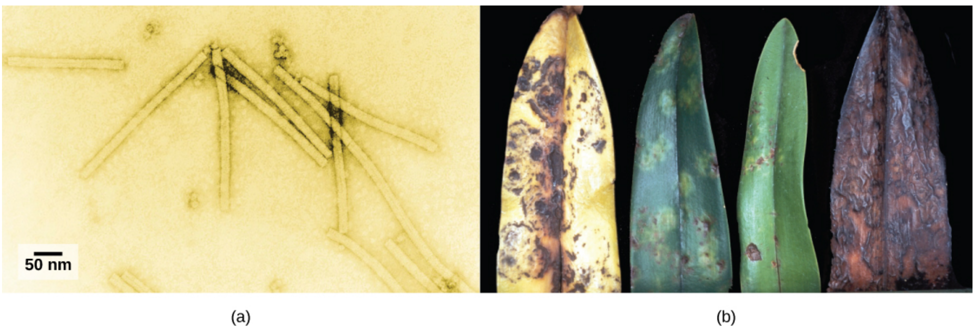
pathogen recognition and memory mechanisms. Research continues to unravel the complexities and vulnerabilities of the immune system.

Despite a poor understanding of the workings of the body in the early 18th century in Europe, the practice of inoculation as a method to prevent the often-deadly effects of smallpox was introduced from the courts of the Ottoman Empire. The method involved causing limited infection with the smallpox virus by introducing the pus of an affected individual to a scratch in an uninfected person. The resulting infection was milder than if it had been caught naturally and mortality rates were shown to be about two percent rather than 30 percent from natural infections. Moreover, the inoculation gave the individual immunity to the disease. It was from these early experiences with inoculation that the methods of vaccination were developed, in which a weakened or relatively harmless (killed) derivative of a pathogen is introduced into the individual. The vaccination induces immunity to the disease with few of the risks of being infected. A modern understanding of the causes of the infectious disease and the mechanisms of the immune system began in the late 19th century and continues to grow today.

Viruses

By the end of this section, you will be able to:

- Describe how viruses were first discovered and how they are detected
- Explain the detailed steps of viral replication
- Describe how vaccines are used in prevention and treatment of viral diseases



(a) The tobacco mosaic virus, seen by transmission electron microscopy, was the first virus to be discovered. (b) The leaves of an infected plant are shown. (credit a: scale-bar data from Matt Russell; credit b: modification of work by USDA, Department of Plant Pathology Archive, North Carolina State University)

No one knows exactly when viruses emerged or from where they came, since viruses do not leave historical footprints such as fossils. Modern viruses are thought to be a mosaic of bits and pieces of nucleic acids picked up from various sources along their respective evolutionary paths. Viruses are **acellular**, parasitic entities that are not classified within any domain because they are not considered alive. They have no plasma membrane, internal organelles, or metabolic processes, and they do not divide. Instead, they infect a host cell and use the host's replication processes to produce progeny virus particles. Viruses infect all forms of organisms including bacteria, archaea, fungi, plants, and animals. Living things grow, metabolize, and reproduce. Viruses replicate, but to do so, they are entirely

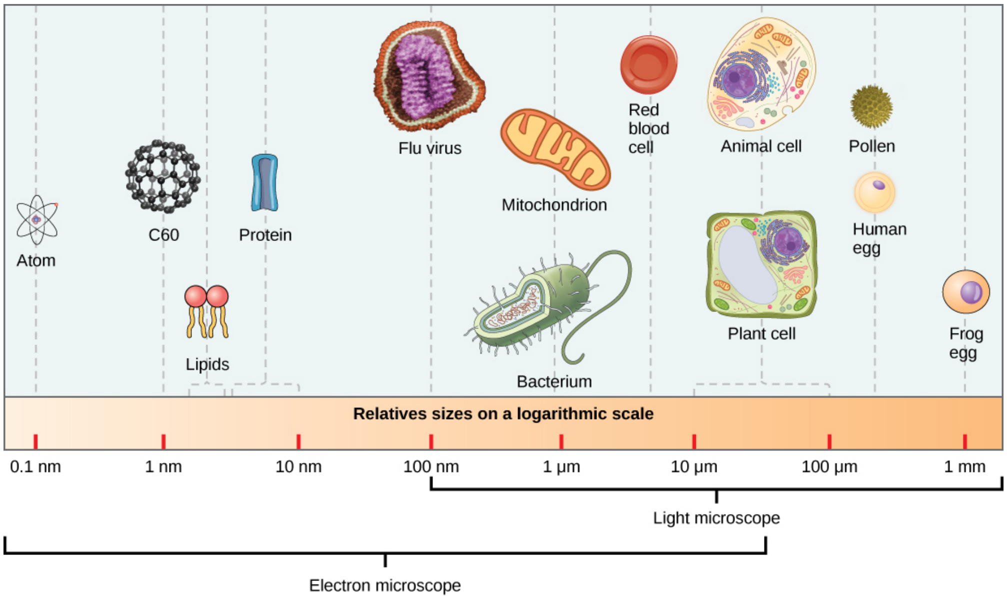
dependent on their host cells. They do not metabolize or grow, but are assembled in their mature form.

Viruses are diverse. They vary in their structure, their replication methods, and in their target hosts or even host cells. While most biological diversity can be understood through evolutionary history, such as how species have adapted to conditions and environments, much about virus origins and evolution remains unknown.

How Viruses Replicate

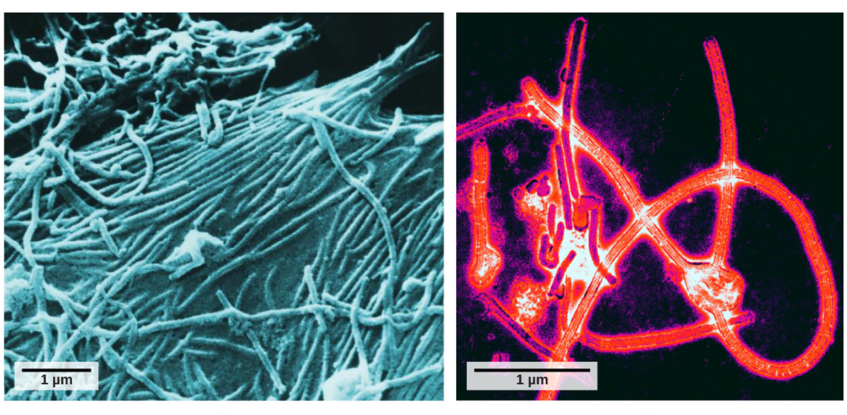
Viruses were first discovered after the development of a porcelain filter, called the Chamberland-Pasteur filter, which could remove all bacteria visible under the microscope from any liquid sample. In 1886, Adolph Meyer demonstrated that a disease of tobacco plants, tobacco mosaic disease, could be transferred from a diseased plant to a healthy one through liquid plant extracts. In 1892, Dmitri Ivanowski showed that this disease could be transmitted in this way even after the Chamberland-Pasteur filter had removed all viable bacteria from the extract. Still, it was many years before it was proven that these “filterable” infectious agents were not simply very small bacteria but were a new type of tiny, disease-causing particle.

Virions, single virus particles, are very small, about 20–250 nanometers (1 nanometer = 1/1,000,000 mm). These individual virus particles are the infectious form of a virus outside the host cell. Unlike bacteria (which are about 100 times larger), we cannot see viruses with a light microscope, with the exception of some large virions of the poxvirus family ([link]).



The size of a virus is very small relative to the size of cells and organelles.

It was not until the development of the electron microscope in the 1940s that scientists got their first good view of the structure of the tobacco mosaic virus ([link]) and others. The surface structure of virions can be observed by both scanning and transmission electron microscopy, whereas the internal structures of the virus can only be observed in images from a transmission electron microscope ([link]).



The ebola virus is shown here as visualized through (a) a scanning electron micrograph and (b) a transmission electron micrograph. (credit a: modification of work by Cynthia Goldsmith, CDC; credit b: modification of work by Thomas W. Geisbert, Boston University School of Medicine; scale-bar data from Matt Russell)

The use of this technology has allowed for the discovery of many viruses of all types of living organisms. They were initially grouped by shared morphology, meaning their size, shape, and distinguishing structures. Later, groups of viruses were classified by the type of nucleic acid they contained, DNA or RNA, and whether their nucleic acid was single- or double-stranded. More recently, molecular analysis of viral replication cycles has further refined their classification.

A **virion** consists of a nucleic-acid core, an outer protein coating, and sometimes an outer envelope made of protein and phospholipid membranes derived from the host cell. The most visible difference between members of viral families is their morphology, which is quite diverse. An interesting feature of viral complexity is that the complexity of the host does not correlate to the complexity of the virion. Some of the most complex virion structures are observed in bacteriophages, viruses that infect the simplest living organisms, bacteria.

Viruses come in many shapes and sizes, but these are consistent and distinct for each viral family ([link]). All virions have a nucleic-acid genome covered by a protective layer of protein, called a **capsid**. The capsid is made of protein subunits called capsomeres. Some viral capsids are simple polyhedral "spheres," whereas others are quite complex in structure. The outer structure surrounding the capsid of some viruses is called the **viral envelope**. All viruses use some sort of **glycoprotein** to attach to their host cells at molecules on the cell called viral receptors. The virus exploits these cell-surface molecules, which the cell uses for some other purpose, as a way

to recognize and infect specific cell types. For example, the measles virus uses a cell-surface glycoprotein in humans that normally functions in immune reactions and possibly in the sperm-egg interaction at fertilization. Attachment is a requirement for viruses to later penetrate the cell membrane, inject the viral genome, and complete their replication inside the cell.

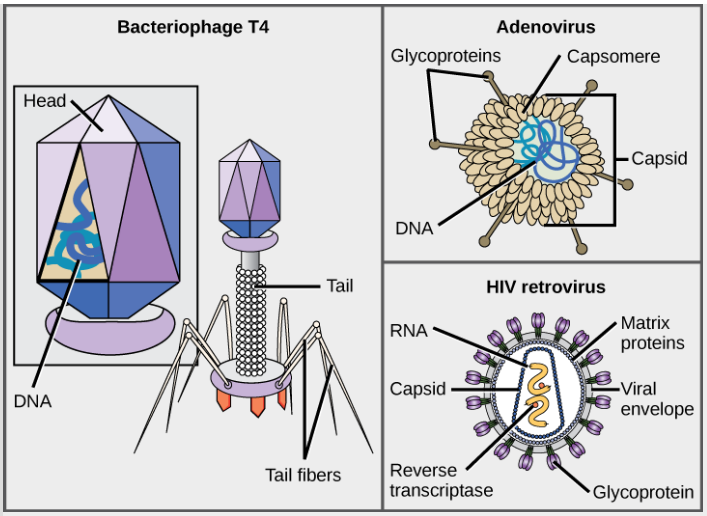
The T4 bacteriophage, which infects the *E. coli* bacterium, is among the most complex virion known; T4 has a protein tail structure that the virus uses to attach to the host cell and a head structure that houses its DNA.

Adenovirus, a nonenveloped animal virus that causes respiratory illnesses in humans, uses protein spikes protruding from its capsomeres to attach to the host cell. Nonenveloped viruses also include those that cause polio (poliovirus), plantar warts (papillomavirus), and hepatitis A (hepatitis A virus). Nonenveloped viruses tend to be more robust and more likely to survive under harsh conditions, such as the gut.

Enveloped virions like HIV (human immunodeficiency virus), the causative agent in AIDS (acquired immune deficiency syndrome), consist of nucleic acid (RNA in the case of HIV) and capsid proteins surrounded by a phospholipid bilayer envelope and its associated proteins ([link]). Chicken pox, influenza, and mumps are examples of diseases caused by viruses with envelopes. Because of the fragility of the envelope, nonenveloped viruses are more resistant to changes in temperature, pH, and some disinfectants than enveloped viruses.

Overall, the shape of the virion and the presence or absence of an envelope tells us little about what diseases the viruses may cause or what species they might infect, but is still a useful means to begin viral classification.

Note:
Art Connection



Viruses can be complex in shape or relatively simple. This figure shows three relatively complex virions: the bacteriophage T4, with its DNA-containing head group and tail fibers that attach to host cells; adenovirus, which uses spikes from its capsid to bind to the host cells; and HIV, which uses glycoproteins embedded in its envelope to do so. Notice that HIV has proteins called matrix proteins, internal to the envelope, which help stabilize virion shape. HIV is a retrovirus, which means it reverse transcribes its RNA genome into DNA, which is then spliced into the host's DNA. (credit "bacteriophage, adenovirus": modification of work by NCBI, NIH; credit "HIV retrovirus": modification of work by NIAID, NIH)

Which of the following statements about virus structure is true?

a. All viruses are encased in a viral membrane.

b. The capsomere is made up of small protein subunits called capsids.
c. DNA is the genetic material in all viruses.
d. Glycoproteins help the virus attach to the host cell.

Unlike all living organisms that use DNA as their genetic material, viruses may use either DNA or RNA as theirs. The virus core contains the genome or total genetic content of the virus. Viral genomes tend to be small compared to bacteria or eukaryotes, containing only those genes that code for proteins the virus cannot get from the host cell. This genetic material may be single-stranded or double-stranded. It may also be linear or circular. While most viruses contain a single segment of nucleic acid, others have genomes that consist of several segments.

DNA viruses have a DNA core. The viral DNA directs the host cell’s replication proteins to synthesize new copies of the viral genome and to transcribe and translate that genome into viral proteins. DNA viruses cause human diseases such as chickenpox, hepatitis B, and some venereal diseases like herpes and genital warts.

RNA viruses contain only RNA in their cores. To replicate their genomes in the host cell, the genomes of RNA viruses encode enzymes not found in host cells. RNA polymerase enzymes are not as stable as DNA polymerases and often make mistakes during transcription. For this reason, mutations, changes in the nucleotide sequence, in RNA viruses occur more frequently than in DNA viruses. This leads to more rapid evolution and change in RNA viruses. For example, the fact that influenza is an RNA virus is one reason a new flu vaccine is needed every year. Human diseases caused by RNA viruses include hepatitis C, measles, and rabies.

Viruses can be seen as obligate intracellular parasites. The virus must attach to a living cell, be taken inside, manufacture its proteins and copy its genome, and find a way to escape the cell so the virus can infect other cells and ultimately other individuals. Viruses can infect only certain species of hosts and only certain cells within that host. The molecular basis for this specificity is that a particular surface molecule, known as the viral receptor,

must be found on the host cell surface for the virus to attach. Also, metabolic differences seen in different cell types based on differential gene expression are a likely factor in which cells a virus may use to replicate. The cell must be making the substances the virus needs, such as enzymes the virus genome itself does not have genes for, or the virus will not be able to replicate using that cell.

Steps of Virus Infections

A virus must “take over” a cell to replicate. The viral replication cycle can produce dramatic biochemical and structural changes in the host cell, which may cause cell damage. These changes, called **cytopathic** effects, can change cell functions or even destroy the cell. Some infected cells, such as those infected by the common cold virus (rhinovirus), die through lysis (bursting) or **apoptosis** (programmed cell death or “cell suicide”), releasing all the progeny virions at once. The symptoms of viral diseases result from the immune response to the virus, which attempts to control and eliminate the virus from the body, and from cell damage caused by the virus. Many animal viruses, such as HIV (human immunodeficiency virus), leave the infected cells of the immune system by a process known as budding, where virions leave the cell individually. During the budding process, the cell does not undergo lysis and is not immediately killed. However, the damage to the cells that HIV infects may make it impossible for the cells to function as mediators of immunity, even though the cells remain alive for a period of time. Most productive viral infections follow similar steps in the virus replication cycle: attachment, penetration, uncoating, replication, assembly, and release.

A virus attaches to a specific receptor site on the host-cell membrane through attachment proteins in the capsid or proteins embedded in its envelope. The attachment is specific, and typically a virus will only attach to cells of one or a few species and only certain cell types within those species with the appropriate receptors.

Note:
Concept in Action

View this [video](#) for a visual explanation of how influenza attacks the body.

Unlike animal viruses, the nucleic acid of bacteriophages is injected into the host cell naked, leaving the capsid outside the cell. Plant and animal viruses can enter their cells through endocytosis, in which the cell membrane surrounds and engulfs the entire virus. Some enveloped viruses enter the cell when the viral envelope fuses directly with the cell membrane. Once inside the cell, the viral capsid is degraded and the viral nucleic acid is released, which then becomes available for replication and transcription.

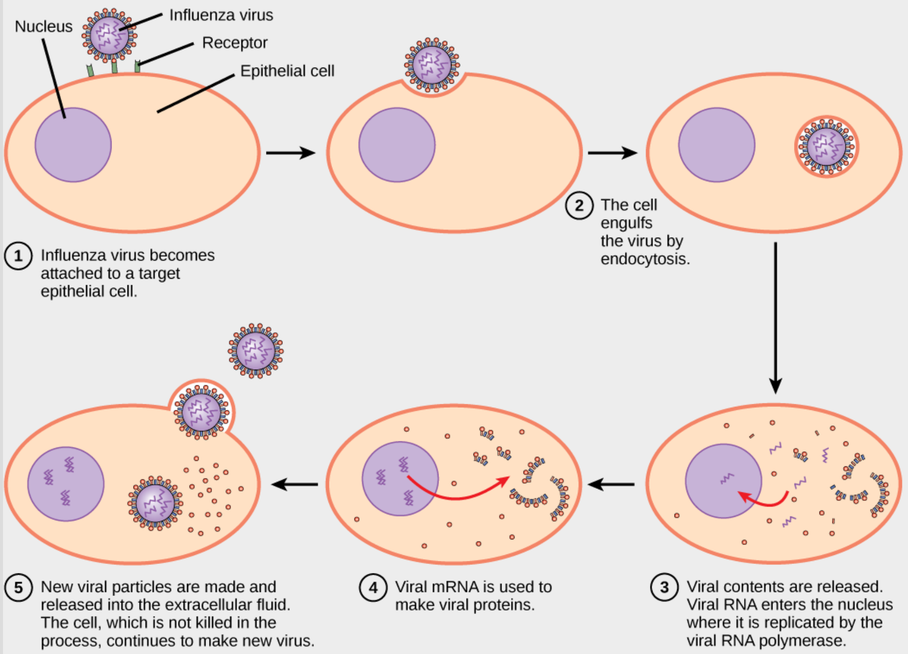
The replication mechanism depends on the viral genome. DNA viruses usually use host cell proteins and enzymes to make additional DNA that is used to copy the genome or be transcribed to messenger RNA (mRNA), which is then used in protein synthesis. RNA viruses, such as the influenza virus, usually use the RNA core as a template for synthesis of viral genomic RNA and mRNA. The viral mRNA is translated into viral enzymes and capsid proteins to assemble new virions ([link]). Of course, there are exceptions to this pattern. If a host cell does not provide the enzymes necessary for viral replication, viral genes supply the information to direct synthesis of the missing proteins. Retroviruses, such as HIV, have an RNA genome that must be reverse transcribed to make DNA, which then is inserted into the host's DNA. To convert RNA into DNA, retroviruses contain genes that encode the virus-specific enzyme reverse transcriptase that transcribes an RNA template to DNA. The fact that HIV produces some of its own enzymes, which are not found in the host, has allowed researchers to develop drugs that inhibit these enzymes. These drugs,

including the reverse transcriptase inhibitor AZT, inhibit HIV replication by reducing the activity of the enzyme without affecting the host's metabolism.

The last stage of viral replication is the release of the new virions into the host organism, where they are able to infect adjacent cells and repeat the replication cycle. Some viruses are released when the host cell dies and other viruses can leave infected cells by budding through the membrane without directly killing the cell.

Note:

Art Connection



In influenza virus infection, glycoproteins attach to a host epithelial cell. As a result, the virus is engulfed. RNA and proteins are made and assembled into new virions.

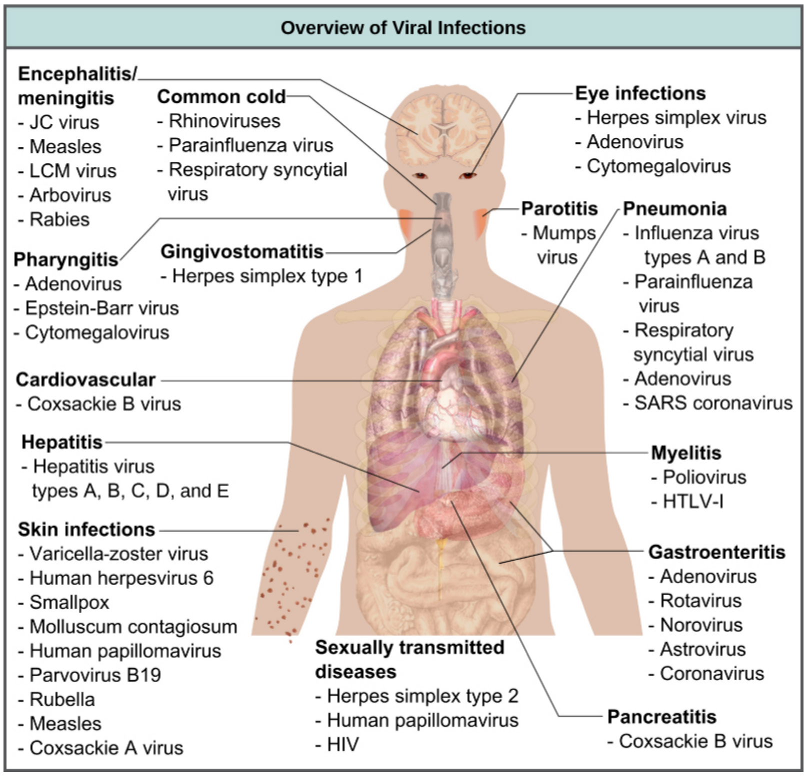
Influenza virus is packaged in a viral envelope, which fuses with the plasma membrane. This way, the virus can exit the host cell without killing it. What advantage does the virus gain by keeping the host cell alive?

Note:
Concept in Action

Click through this tutorial on viruses to identify structures, modes of transmission, replication, and more.

Viruses and Disease

Viruses cause a variety of diseases in animals, including humans, ranging from the common cold to potentially fatal illnesses like meningitis ([link]). These diseases can be treated by antiviral drugs or by vaccines, but some viruses, such as HIV, are capable of avoiding the immune response and mutating so as to become resistant to antiviral drugs.



Viruses are the cause of dozens of ailments in humans, ranging from mild illnesses to serious diseases. (credit: modification of work by Mikael Häggström)

Vaccines for Prevention

While we do have limited numbers of effective antiviral drugs, such as those used to treat HIV and influenza, the primary method of controlling viral disease is by vaccination, which is intended to prevent outbreaks by building immunity to a virus or virus family. A **vaccine** may be prepared using weakened live viruses, killed viruses, or molecular subunits of the virus. In general, live viruses lead to better immunity, but have the possibility of causing disease at some low frequency. Killed viral vaccine and the subunit viruses are both incapable of causing disease, but in general lead to less effective or long-lasting immunity.

Weakened live viral vaccines are designed in the laboratory to cause few symptoms in recipients while giving them immunity against future infections. Polio was one disease that represented a milestone in the use of vaccines. Mass immunization campaigns in the U.S. in the 1950s (killed vaccine) and 1960s (live vaccine) essentially eradicated the disease, which caused muscle paralysis in children and generated fear in the general population when regional epidemics occurred. The success of the polio vaccine paved the way for the routine dispensation of childhood vaccines against measles, mumps, rubella, chickenpox, and other diseases.

Live vaccines are usually made by **attenuation** (weakening) of the “wild-type” (disease-causing) virus by growing it in the laboratory in tissues or at temperatures different from what the virus is accustomed to in the host. For example, the virus may be grown in cells in a test tube, in bird embryos, or in live animals. The adaptation to these new cells or temperature induces mutations in the virus’ genomes, allowing them to grow better in the laboratory while inhibiting their ability to cause disease when reintroduced into the conditions found in the host. These attenuated viruses thus still cause an infection, but they do not grow very well, allowing the immune response to develop in time to prevent major disease. The danger of using live vaccines, which are usually more effective than killed vaccines, is the low but significant risk that these viruses will revert back to their disease-causing form by back mutations. Back mutations occur when the vaccine undergoes mutations in the host such that it readapts to the host and can again cause disease, which can then be spread to other humans in an epidemic. This happened as recently as 2007 in Nigeria where mutations in a polio vaccine led to an epidemic of polio in that country.

Some vaccines are in continuous development because certain viruses, such as influenza and HIV, have a high mutation rate compared to other viruses or host cells. With influenza, mutation in genes for the surface molecules helps the virus evade the protective immunity that may have been obtained in a previous influenza season, making it necessary for individuals to get vaccinated every year. Other viruses, such as those that cause the childhood diseases measles, mumps, and rubella, mutate so little that the same vaccine is used year after year.

Vaccines and Antiviral Drugs for Treatment

In some cases, vaccines can be used to treat an active viral infection. In the case of rabies, a fatal neurological disease transmitted in the saliva of rabies virus-infected animals, the progression of the disease from the time of the animal bite to the time it enters the central nervous system may be two weeks or longer. This is enough time to vaccinate an individual who suspects being bitten by a rabid animal, and the boosted immune response from the vaccination is enough to prevent the virus from entering nervous tissue. Thus, the fatal neurological consequences of the disease are averted and the individual only has to recover from the infected bite. This approach is also being used for the treatment of Ebola, one of the fastest and most deadly viruses affecting humans, though usually infecting limited populations. Ebola is also a leading cause of death in gorillas. Transmitted by bats and great apes, this virus can cause death in 70–90 percent of the infected within two weeks. Using newly developed vaccines that boost the immune response, there is hope that immune systems of affected individuals will be better able to control the virus, potentially reducing mortality rates.

Another way of treating viral infections is the use of antiviral drugs. These drugs often have limited ability to cure viral disease but have been used to control and reduce symptoms for a wide variety of viral diseases. For most viruses, these drugs inhibit the virus by blocking the actions of one or more of its proteins. It is important that the targeted proteins be encoded for by viral genes and that these molecules are not present in a healthy host cell. In this way, viral growth is inhibited without damaging the host. There are large numbers of antiviral drugs available to treat infections, some specific for a particular virus and others that can affect multiple viruses.

Antivirals have been developed to treat genital herpes (herpes simplex II) and influenza. For genital herpes, drugs such as acyclovir can reduce the number and duration of the episodes of active viral disease during which patients develop viral lesions in their skins cells. As the virus remains latent in nervous tissue of the body for life, this drug is not a cure but can make the symptoms of the disease more manageable. For influenza, drugs like Tamiflu can reduce the duration of “flu” symptoms by one or two days, but

the drug does not prevent symptoms entirely. Other antiviral drugs, such as Ribavirin, have been used to treat a variety of viral infections.

By far the most successful use of antivirals has been in the treatment of the retrovirus HIV, which causes a disease that, if untreated, is usually fatal within 10–12 years after being infected. Anti-HIV drugs have been able to control viral replication to the point that individuals receiving these drugs survive for a significantly longer time than the untreated.

Anti-HIV drugs inhibit viral replication at many different phases of the HIV replicative cycle. Drugs have been developed that inhibit the fusion of the HIV viral envelope with the plasma membrane of the host cell (fusion inhibitors), the conversion of its RNA genome to double-stranded DNA (reverse transcriptase inhibitors), the integration of the viral DNA into the host genome (integrase inhibitors), and the processing of viral proteins (protease inhibitors).

When any of these drugs are used individually, the virus' high mutation rate allows the virus to rapidly evolve resistance to the drug. The breakthrough in the treatment of HIV was the development of highly active anti-retroviral therapy (HAART), which involves a mixture of different drugs, sometimes called a drug "cocktail." By attacking the virus at different stages of its replication cycle, it is difficult for the virus to develop resistance to multiple drugs at the same time. Still, even with the use of combination HAART therapy, there is concern that, over time, the virus will evolve resistance to this therapy. Thus, new anti-HIV drugs are constantly being developed with the hope of continuing the battle against this highly fatal virus.

Section Summary

Viruses are acellular entities that can usually only be seen with an electron microscope. Their genomes contain either DNA or RNA, and they replicate using the replication proteins of a host cell. Viruses are diverse, infecting archaea, bacteria, fungi, plants, and animals. Viruses consist of a nucleic-acid core surrounded by a protein capsid with or without an outer lipid envelope.

Viral replication within a living cell always produces changes in the cell, sometimes resulting in cell death and sometimes slowly killing the infected cells. There are six basic stages in the virus replication cycle: attachment, penetration, uncoating, replication, assembly, and release. A viral infection may be productive, resulting in new virions, or nonproductive, meaning the virus remains inside the cell without producing new virions.

Viruses cause a variety of diseases in humans. Many of these diseases can be prevented by the use of viral vaccines, which stimulate protective immunity against the virus without causing major disease. Viral vaccines may also be used in active viral infections, boosting the ability of the immune system to control or destroy the virus. Antiviral drugs that target enzymes and other protein products of viral genes have been developed and used with mixed success. Combinations of anti-HIV drugs have been used to effectively control the virus, extending the lifespan of infected individuals.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about virus structure is true?

a. All viruses are encased in a viral membrane.
b. The capsomere is made up of small protein subunits called capsids.
c. DNA is the genetic material in all viruses.
d. Glycoproteins help the virus attach to the host cell.

Solution:

[link] D

Exercise:

Problem:

[link] Influenza virus is packaged in a viral envelope, which fuses with the plasma membrane. This way, the virus can exit the host cell without killing it. What advantage does the virus gain by keeping the host cell alive?

Solution:

[link] The host cell can continue to make new virus particles.

Review Questions

Exercise:

Problem: Which statement is true?

a. A virion contains DNA and RNA.
b. Viruses are acellular.
c. Viruses replicate outside of the cell.
d. Most viruses are easily visualized with a light microscope.

Solution:

B

Exercise:

Problem:

The viral _________ plays a role in attaching a virion to the host cell.

a. core
b. capsid
c. envelope
d. both b and c

Solution:

D

Exercise:

Problem: Which statement is true of viral replication?

a. In the process of apoptosis, the cell survives.
b. During attachment, the virus attaches at specific sites on the cell surface.
c. The viral capsid helps the host cell produce more copies of the viral genome.
d. mRNA works outside of the host cell to produce enzymes and proteins.

Solution:

B

Free Response

Exercise:

Problem: Why can’t dogs catch the measles?

Solution:

The virus cannot attach to dog cells because dog cells do not express the receptors for the virus or there is no cell within the dog that is permissive for viral replication.

Exercise:

Problem:

Why is immunization after being bitten by a rabid animal so effective?

Solution:

Rabies vaccine works after a bite because it takes two weeks for the virus to travel from the site of the bite to the central nervous system, where the most severe symptoms of the disease occur. The vaccine is able to cause an immune response in the body during this time that clears the infection before it reaches the nervous system.

Glossary

acellular
: lacking cells

apoptosis
: the cell death caused by induction of a cell's own internal mechanisms either as a natural step in the development of a multicellular organism or by other environmental factors such as signals from cells of the immune system

attenuation
: the weakening of a virus during vaccine development

capsid
: the protein coating of the viral core

cytopathic
: causing cell damage

glycoprotein
: a protein molecule with attached carbohydrate molecules

vaccine

a weakened solution of virus components, viruses, or other agents that produce an immune response

virion
an individual virus particle outside a host cell

viral envelope
a lipid bilayer that envelops some viruses

Innate Immunity

By the end of this section, you will be able to:

- Describe the body’s innate physical and chemical defenses
- Explain the inflammatory response
- Describe the complement system

The vertebrate, including human, immune system is a complex multilayered system for defending against external and internal threats to the integrity of the body. The system can be divided into two types of defense systems: the innate immune system, which is nonspecific toward a particular kind of pathogen, and the adaptive immune system, which is specific ([link]). **Innate immunity** is not caused by an infection or vaccination and depends initially on physical and chemical barriers that work on all pathogens, sometimes called the first line of defense. The second line of defense of the innate system includes chemical signals that produce inflammation and fever responses as well as mobilizing protective cells and other chemical defenses. The adaptive immune system mounts a highly specific response to substances and organisms that do not belong in the body. The adaptive system takes longer to respond and has a memory system that allows it to respond with greater intensity should the body reencounter a pathogen even years later.

Vertebrate Immunity		
Innate Immune System		Adaptive Immune System
Physical Barriers	Internal Defenses	
• Skin, hair, cilia • Mucus membranes • Mucus and chemical secretions • Digestive enzymes in mouth • Stomach acid	• Inflammatory response • Complement proteins • Phagocytic cells • Natural killer (NK) cells	• Antibodies and the humoral immune response • Cell-mediated immune response • Memory response

There are two main parts to the vertebrate immune system. The innate immune system, which is made up of physical barriers and internal defenses, responds to all pathogens. The adaptive immune system is highly specific.

External and Chemical Barriers

The body has significant physical barriers to potential pathogens. The skin contains the protein keratin, which resists physical entry into cells. Other body surfaces, particularly those associated with body openings, are protected by the mucous membranes. The sticky mucus provides a physical trap for pathogens, preventing their movement deeper into the body. The openings of the body, such as the nose and ears, are protected by hairs that catch pathogens, and the mucous membranes of the upper respiratory tract have cilia that constantly move pathogens trapped in the mucus coat up to the mouth.

The skin and mucous membranes also create a chemical environment that is hostile to many microorganisms. The surface of the skin is acidic, which prevents bacterial growth. Saliva, mucus, and the tears of the eye contain an enzyme that breaks down bacterial cell walls. The stomach secretions create a highly acidic environment, which kills many pathogens entering the digestive system.

Finally, the surface of the body and the lower digestive system have a community of microorganisms such as bacteria, archaea, and fungi that coexist without harming the body. There is evidence that these organisms are highly beneficial to their host, combating disease-causing organisms and outcompeting them for nutritional resources provided by the host body. Despite these defenses, pathogens may enter the body through skin abrasions or punctures, or by collecting on mucosal surfaces in large numbers that overcome the protections of mucus or cilia.

Internal Defenses

When pathogens enter the body, the innate immune system responds with a variety of internal defenses. These include the inflammatory response, phagocytosis, natural killer cells, and the complement system. White blood cells in the blood and lymph recognize pathogens as foreign to the body. A **white blood cell** is larger than a red blood cell, is nucleated, and is typically able to move using amoeboid locomotion. Because they can move on their own, white blood cells can leave the blood to go to infected tissues. For

example, a **monocyte** is a type of white blood cell that circulates in the blood and lymph and develops into a macrophage after it moves into infected tissue. A **macrophage** is a large cell that engulfs foreign particles and pathogens. **Mast cells** are produced in the same way as white blood cells, but unlike circulating white blood cells, mast cells take up residence in connective tissues and especially mucosal tissues. They are responsible for releasing chemicals in response to physical injury. They also play a role in the allergic response, which will be discussed later in the chapter.

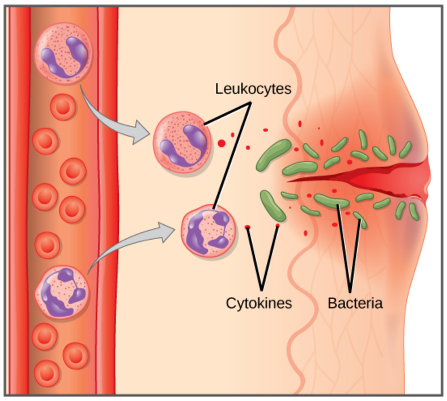
When a pathogen is recognized as foreign, chemicals called cytokines are released. A **cytokine** is a chemical messenger that regulates cell differentiation (form and function), proliferation (production), and gene expression to produce a variety of immune responses. Approximately 40 types of cytokines exist in humans. In addition to being released from white blood cells after pathogen recognition, cytokines are also released by the infected cells and bind to nearby uninfected cells, inducing those cells to release cytokines. This positive feedback loop results in a burst of cytokine production.

One class of early-acting cytokines is the interferons, which are released by infected cells as a warning to nearby uninfected cells. An **interferon** is a small protein that signals a viral infection to other cells. The interferons stimulate uninfected cells to produce compounds that interfere with viral replication. Interferons also activate macrophages and other cells.

The Inflammatory Response and Phagocytosis

The first cytokines to be produced encourage **inflammation**, a localized redness, swelling, heat, and pain. Inflammation is a response to physical trauma, such as a cut or a blow, chemical irritation, and infection by pathogens (viruses, bacteria, or fungi). The chemical signals that trigger an inflammatory response enter the extracellular fluid and cause capillaries to dilate (expand) and capillary walls to become more permeable, or leaky. The serum and other compounds leaking from capillaries cause swelling of the area, which in turn causes pain. Various kinds of white blood cells are attracted to the area of inflammation. The types of white blood cells that

arrive at an inflamed site depend on the nature of the injury or infecting pathogen. For example, a **neutrophil** is an early arriving white blood cell that engulfs and digests pathogens. Neutrophils are the most abundant white blood cells of the immune system ([link]). Macrophages follow neutrophils and take over the phagocytosis function and are involved in the resolution of an inflamed site, cleaning up cell debris and pathogens.



White blood cells (leukocytes) release chemicals to stimulate the inflammatory response following a cut in the skin.

Cytokines also send feedback to cells of the nervous system to bring about the overall symptoms of feeling sick, which include lethargy, muscle pain, and nausea. Cytokines also increase the core body temperature, causing a fever. The elevated temperatures of a fever inhibit the growth of pathogens and speed up cellular repair processes. For these reasons, suppression of fevers should be limited to those that are dangerously high.

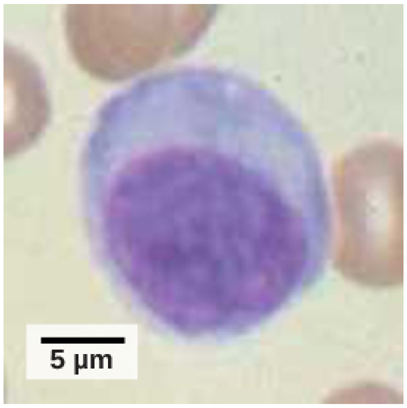
Note:

Concept in Action

Check out this 23-second, stop-motion video showing a neutrophil that searches and engulfs fungus spores during an elapsed time of 79 minutes.

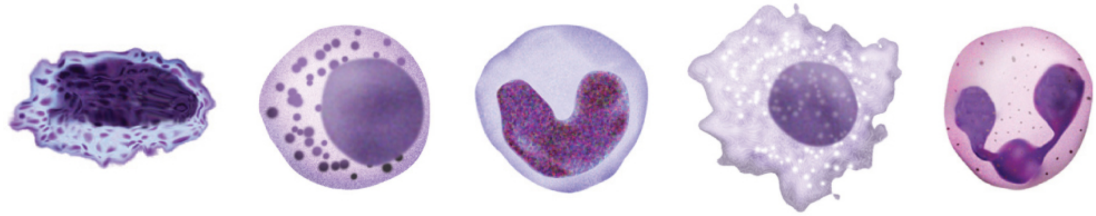
Natural Killer Cells

A **lymphocyte** is a white blood cell that contains a large nucleus ([link]). Most lymphocytes are associated with the adaptive immune response, but infected cells are identified and destroyed by natural killer cells, the only lymphocytes of the innate immune system. A **natural killer (NK) cell** is a lymphocyte that can kill cells infected with viruses (or cancerous cells). NK cells identify intracellular infections, especially from viruses, by the altered expression of **major histocompatibility class (MHC) I molecules** on the surface of infected cells. MHC class I molecules are proteins on the surfaces of all nucleated cells that provide a sample of the cell's internal environment at any given time. Unhealthy cells, whether infected or cancerous, display an altered MHC class I complement on their cell surfaces.



Lymphocytes, such as NK cells, are characterized by their large nuclei that actively absorb Wright stain and therefore appear dark colored under a microscope. (credit: scale-bar data from Matt Russell)

After the NK cell detects an infected or tumor cell, it induces programmed cell death, or apoptosis. Phagocytic cells then come along and digest the cell debris left behind. NK cells are constantly patrolling the body and are an effective mechanism for controlling potential infections and preventing cancer progression. The various types of immune cells are shown in [link].



Cells involved in the innate immune response include mast cells, natural killer cells, and white blood cells, such as monocytes, macrophages and neutrophils.

Complement

An array of approximately 20 types of proteins, called a **complement system**, is also activated by infection or the activity of the cells of the adaptive immune system and functions to destroy extracellular pathogens. Liver cells and macrophages synthesize inactive forms of complement proteins continuously; these proteins are abundant in the blood serum and are capable of responding immediately to infecting microorganisms. The complement system is so named because it is complementary to the innate and adaptive immune system. Complement proteins bind to the surfaces of microorganisms and are particularly attracted to pathogens that are already tagged by the adaptive immune system. This “tagging” involves the attachment of specific proteins called antibodies (discussed in detail later) to the pathogen. When they attach, the antibodies change shape providing a binding site for one of the complement proteins. After the first few complement proteins bind, a cascade of binding in a specific sequence of proteins follows in which the pathogen rapidly becomes coated in complement proteins.

Complement proteins perform several functions, one of which is to serve as a marker to indicate the presence of a pathogen to phagocytic cells and enhance engulfment. Certain complement proteins can combine to open pores in microbial cell membranes and cause lysis of the cells.

Section Summary

The innate immune system consists first of physical and chemical barriers to infection including the skin and mucous membranes and their secretions, ciliated surfaces, and body hairs. The second line of defense is an internal defense system designed to counter pathogenic threats that bypass the physical and chemical barriers of the body. Using a combination of cellular and molecular responses, the innate immune system identifies the nature of a pathogen and responds with inflammation, phagocytosis, cytokine release, destruction by NK cells, or the complement system.

Review Questions

Exercise:

Problem:

Which of the following is a barrier against pathogens provided by the skin?

a. low pH
b. mucus
c. tears
d. cilia

Solution:

A

Exercise:

Problem:

Although interferons have several effects, they are particularly useful against infections with which type of pathogen?

a. bacteria
b. viruses

c. fungi
d. helminths

Solution:

B

Exercise:

Problem:

Which innate immune system component uses MHC class I molecules directly in its defense strategy?

a. macrophages
b. neutrophils
c. NK cells
d. interferon

Solution:

C

Free Response

Exercise:

Problem:

Different MHC class I molecules between donor and recipient cells can lead to rejection of a transplanted organ or tissue. Suggest a reason for this.

Solution:

If the MHC class I molecules expressed on donor cells differ from the MHC class I molecules expressed on recipient cells, NK cells may

identify the donor cells as not normal and produce enzymes to induce the donor cells to undergo apoptosis, which would destroy the transplanted organ.

Exercise:

Problem:

If a series of genetic mutations prevented some, but not all, of the complement proteins from binding antibodies or pathogens, would the entire complement system be compromised?

Solution:

The entire complement system would probably be affected even when only a few members were mutated such that they could no longer bind. Because the complement involves the binding of activated proteins in a specific sequence, when one or more proteins in the sequence is absent, the subsequent proteins would be incapable of binding to elicit the complement's pathogen-destructive effects.

Glossary

complement system
: an array of approximately 20 soluble proteins of the innate immune system that enhance phagocytosis, bore holes in pathogens, and recruit lymphocytes

cytokine
: a chemical messenger that regulates cell differentiation, proliferation, and gene expression to effect immune responses

inflammation
: the localized redness, swelling, heat, and pain that results from the movement of leukocytes through opened capillaries to a site of infection

innate immunity

an immunity that occurs naturally because of genetic factors or physiology, and is not caused by infection or vaccination

interferon
: a cytokine that inhibits viral replication

lymphocyte
: a type of white blood cell that includes natural killer cells of the innate immune system and B and T cells of the adaptive immune system

macrophage
: a large phagocytic cell that engulfs foreign particles and pathogens

major histocompatibility class (MHC) I
: a group of proteins found on the surface of all nucleated cells that signals to immune cells whether the cell is normal or is infected or cancerous; it also provides the appropriate sites into which antigens can be loaded for recognition by lymphocytes

mast cell
: a leukocyte that produces inflammatory molecules, such as histamine, in response to large pathogens

monocyte
: a type of white blood cell that circulates in the blood and lymph and differentiates into a macrophage after it moves into infected tissue

natural killer (NK) cell
: a lymphocyte that can kill cells infected with viruses or tumor cells

neutrophil
: a phagocytic leukocyte that engulfs and digests pathogens

white blood cell
: a nucleated cell found in the blood that is a part of the immune system; also called leukocytes

Adaptive Immunity
By the end of this section, you will be able to:

- Explain adaptive immunity
- Describe cell-mediated immune response and humoral immune response
- Describe immune tolerance

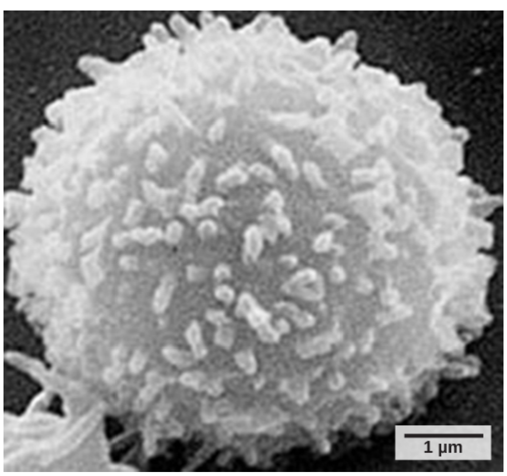
The adaptive, or acquired, immune response takes days or even weeks to become established—much longer than the innate response; however, adaptive immunity is more specific to an invading pathogen. **Adaptive immunity** is an immunity that occurs after exposure to an antigen either from a pathogen or a vaccination. An **antigen** is a molecule that stimulates a response in the immune system. This part of the immune system is activated when the innate immune response is insufficient to control an infection. In fact, without information from the innate immune system, the adaptive response could not be mobilized. There are two types of adaptive responses: the **cell-mediated immune response**, which is controlled by activated **T cells**, and the **humoral immune response**, which is controlled by activated **B cells** and antibodies. Activated T and B cells whose surface binding sites are specific to the molecules on the pathogen greatly increase in numbers and attack the invading pathogen. Their attack can kill pathogens directly or they can secrete antibodies that enhance the phagocytosis of pathogens and disrupt the infection. Adaptive immunity also involves a memory to give the host long-term protection from reinfection with the same type of pathogen; on reexposure, this host memory will facilitate a rapid and powerful response.

B and T Cells

Lymphocytes, which are white blood cells, are formed with other blood cells in the red bone marrow found in many flat bones, such as the shoulder or pelvic bones. The two types of lymphocytes of the adaptive immune response are B and T cells ([link]). Whether an immature lymphocyte becomes a B cell or T cell depends on where in the body it matures. The B cells remain in the bone marrow to mature (hence the name “B” for “bone

marrow”), while T cells migrate to the thymus, where they mature (hence the name “T” for “thymus”).

Maturation of a B or T cell involves becoming immunocompetent, meaning that it can recognize, by binding, a specific molecule or antigen (discussed below). During the maturation process, B and T cells that bind too strongly to the body’s own cells are eliminated in order to minimize an immune response against the body’s own tissues. Those cells that react weakly to the body’s own cells, but have highly specific receptors on their cell surfaces that allow them to recognize a foreign molecule, or antigen, remain. This process occurs during fetal development and continues throughout life. The specificity of this receptor is determined by the genetics of the individual and is present before a foreign molecule is introduced to the body or encountered. Thus, it is genetics and not experience that initially provides a vast array of cells, each capable of binding to a different specific foreign molecule. Once they are immunocompetent, the T and B cells will migrate to the spleen and lymph nodes where they will remain until they are called on during an infection. B cells are involved in the humoral immune response, which targets pathogens loose in blood and lymph, and T cells are involved in the cell-mediated immune response, which targets infected cells.

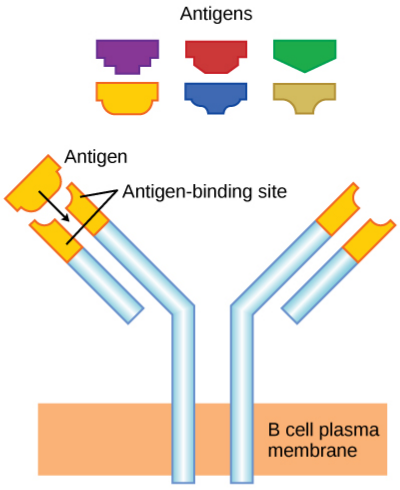


This scanning electron micrograph shows a T

lymphocyte. T and B cells are indistinguishable by light microscopy but can be differentiated experimentally by probing their surface receptors. (credit: modification of work by NCI; scale-bar data from Matt Russell)

Humoral Immune Response

As mentioned, an antigen is a molecule that stimulates a response in the immune system. Not every molecule is antigenic. B cells participate in a chemical response to antigens present in the body by producing specific antibodies that circulate throughout the body and bind with the antigen whenever it is encountered. This is known as the humoral immune response. As discussed, during maturation of B cells, a set of highly specific B cells are produced that have many antigen receptor molecules in their membrane ([link]).



B cell receptors are embedded in the membranes of B cells and bind a variety of antigens through their variable regions.

Each B cell has only one kind of antigen receptor, which makes every B cell different. Once the B cells mature in the bone marrow, they migrate to lymph nodes or other lymphatic organs. When a B cell encounters the antigen that binds to its receptor, the antigen molecule is brought into the cell by endocytosis and reappears on the surface of the cell bound to an **MHC class II molecule**. When this process is complete, the B cell is sensitized. In most cases, the sensitized B cell must then encounter a specific kind of T cell, called a helper T cell, before it is activated. The helper T cell must already have been activated through an encounter with the antigen (discussed below).

The helper T cell binds to the antigen-MHC class II complex and is induced to release cytokines that induce the B cell to divide rapidly, which makes thousands of identical (clonal) cells. These daughter cells become either plasma cells or memory B cells. The memory B cells remain inactive at this point, until another later encounter with the antigen, caused by a reinfection by the same bacteria or virus, results in them dividing into a new population of plasma cells. The plasma cells, on the other hand, produce and secrete large quantities, up to 100 million molecules per hour, of antibody molecules. An **antibody**, also known as an immunoglobulin (Ig), is a protein that is produced by plasma cells after stimulation by an antigen. Antibodies are the agents of humoral immunity. Antibodies occur in the blood, in gastric and mucus secretions, and in breast milk. Antibodies in these bodily fluids can bind pathogens and mark them for destruction by phagocytes before they can infect cells.

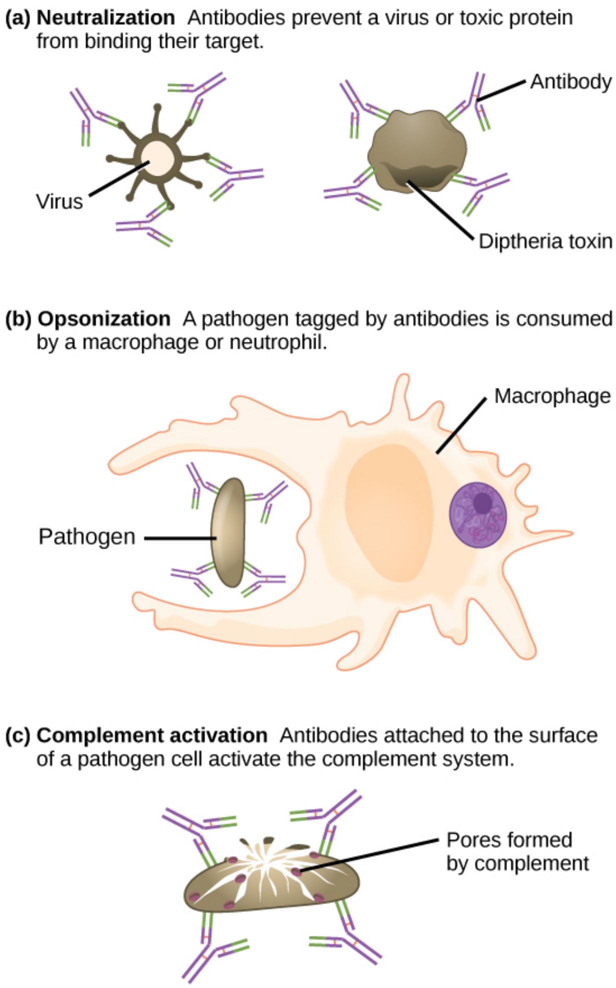
These antibodies circulate in the blood stream and lymphatic system and bind with the antigen whenever it is encountered. The binding can fight infection in several ways. Antibodies can bind to viruses or bacteria and interfere with the chemical interactions required for them to infect or bind to other cells. The antibodies may create bridges between different particles containing antigenic sites clumping them all together and preventing their proper functioning. The antigen-antibody complex stimulates the complement system described previously, destroying the cell bearing the antigen. Phagocytic cells, such as those already described, are attracted by the antigen-antibody complexes, and phagocytosis is enhanced when the complexes are present. Finally, antibodies stimulate inflammation, and their presence in mucus and on the skin prevents pathogen attack.

Antibodies coat extracellular pathogens and neutralize them by blocking key sites on the pathogen that enhance their infectivity (such as receptors that “dock” pathogens on host cells) ([link]). Antibody neutralization can prevent pathogens from entering and infecting host cells. The neutralized antibody-coated pathogens can then be filtered by the spleen and eliminated in urine or feces.

Antibodies also mark pathogens for destruction by phagocytic cells, such as macrophages or neutrophils, in a process called opsonization. In a process called complement fixation, some antibodies provide a place for complement proteins to bind. The combination of antibodies and complement promotes rapid clearing of pathogens.

The production of antibodies by plasma cells in response to an antigen is called **active immunity** and describes the host’s active response of the immune system to an infection or to a vaccination. There is also a **passive immune** response where antibodies come from an outside source, instead of the individual’s own plasma cells, and are introduced into the host. For example, antibodies circulating in a pregnant woman’s body move across the placenta into the developing fetus. The child benefits from the presence of these antibodies for up to several months after birth. In addition, a passive immune response is possible by injecting antibodies into an individual in the form of an antivenom to a snake-bite toxin or antibodies in blood serum to help fight a hepatitis infection. This gives immediate

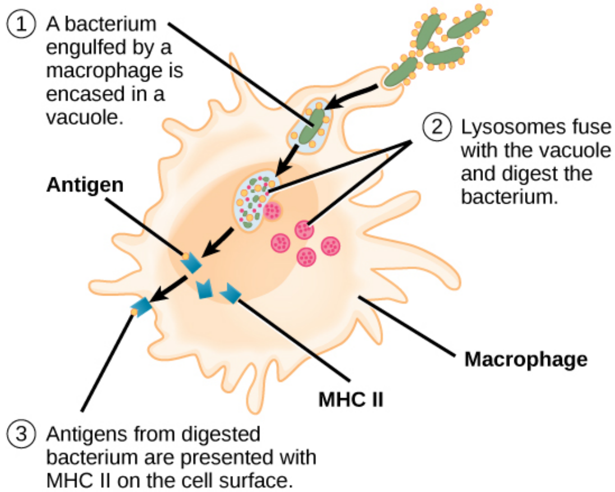
protection since the body does not need the time required to mount its own response.



Antibodies may inhibit infection by (a) preventing the antigen from binding its target, (b) tagging a pathogen for destruction by macrophages or neutrophils, or (c) activating the complement cascade.

Cell-Mediated Immunity

Unlike B cells, T lymphocytes are unable to recognize pathogens without assistance. Instead, dendritic cells and macrophages first engulf and digest pathogens into hundreds or thousands of antigens. Then, an **antigen-presenting cell (APC)** detects, engulfs, and informs the adaptive immune response about an infection. When a pathogen is detected, these APCs will engulf and break it down through phagocytosis. Antigen fragments will then be transported to the surface of the APC, where they will serve as an indicator to other immune cells. A **dendritic cell** is an immune cell that mops up antigenic materials in its surroundings and presents them on its surface. Dendritic cells are located in the skin, the linings of the nose, lungs, stomach, and intestines. These positions are ideal locations to encounter invading pathogens. Once they are activated by pathogens and mature to become APCs they migrate to the spleen or a lymph node. Macrophages also function as APCs. After phagocytosis by a macrophage, the phagocytic vesicle fuses with an intracellular lysosome. Within the resulting phagolysosome, the components are broken down into fragments; the fragments are then loaded onto MHC class II molecules and are transported to the cell surface for antigen presentation ([link]). Helper T cells cannot properly respond to an antigen unless it is processed and embedded in an MHC class II molecule. The APCs express MHC class II on their surfaces, and when combined with a foreign antigen, these complexes signal an invader.



An antigen-presenting cell (APC), such as a macrophage, engulfs a foreign antigen, partially digests it in a lysosome, and then embeds it in an MHC class II molecule for presentation at the cell surface. Lymphocytes of the adaptive immune response must interact with antigen-embedded MHC class II molecules to mature into functional immune cells.

Note:
Concept in Action

View this animation from Rockefeller University to see how dendritic cells act as sentinels in the body’s immune system.

T cells have many functions. Some respond to APCs of the innate immune system and indirectly induce immune responses by releasing cytokines. Others stimulate B cells to start the humoral response as described previously. Another type of T cell detects APC signals and directly kills the infected cells, while some are involved in suppressing inappropriate immune reactions to harmless or “self” antigens.

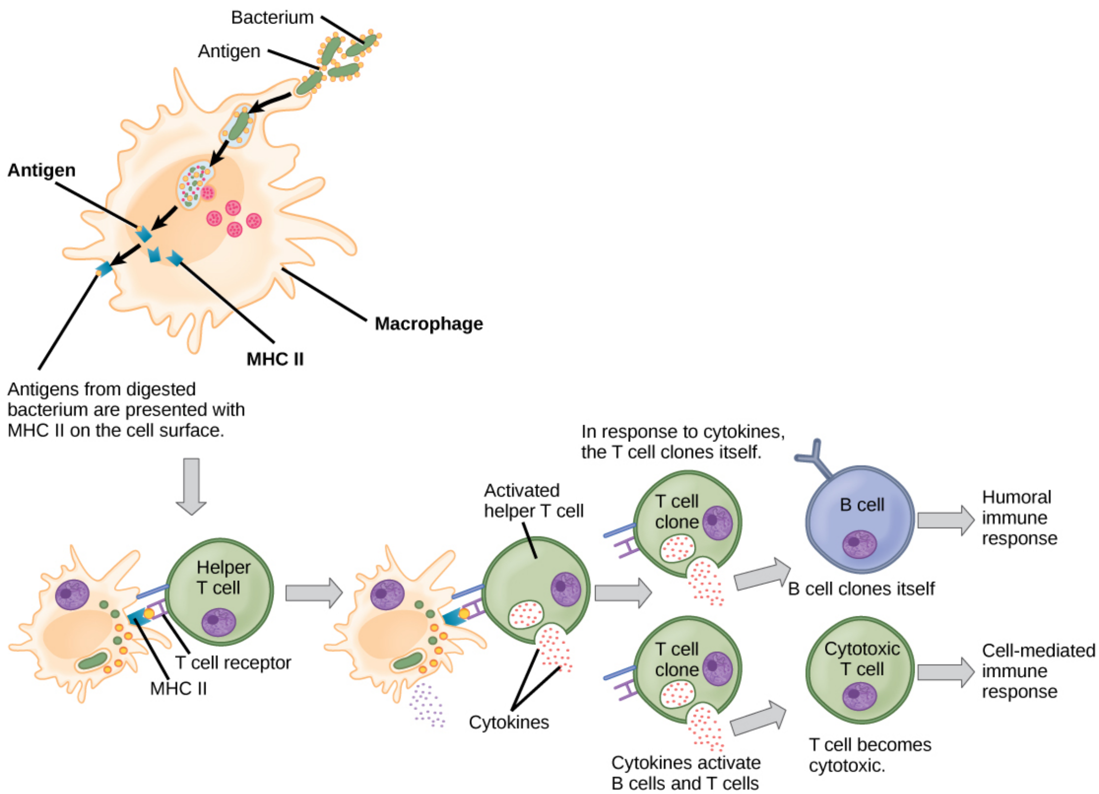
There are two main types of T cells: helper T lymphocytes (T_H) and the cytotoxic T lymphocytes (T_C). The T_H lymphocytes function indirectly to tell other immune cells about potential pathogens. T_H lymphocytes recognize specific antigens presented by the MHC class II complexes of APCs. There are two populations of T_H cells: T_H1 and T_H2. T_H1 cells secrete cytokines to enhance the activities of macrophages and other T cells. T_H2 cells stimulate naïve B cells to secrete antibodies. Whether a T_H1 or a T_H2 immune response develops depends on the specific types of cytokines secreted by cells of the innate immune system, which in turn depends on the nature of the invading pathogen.

Cytotoxic T cells (T_C) are the key component of the cell-mediated part of the adaptive immune system and attack and destroy infected cells. T_C cells are particularly important in protecting against viral infections; this is because viruses replicate within cells where they are shielded from extracellular contact with circulating antibodies. Once activated, the T_C creates a large clone of cells with one specific set of cell-surface receptors, as in the case with proliferation of activated B cells. As with B cells, the clone includes active T_C cells and inactive memory T_C cells. The resulting active T_C cells then identify infected host cells. Because of the time required to generate a population of clonal T and B cells, there is a delay in the adaptive immune response compared to the innate immune response.

T_C cells attempt to identify and destroy infected cells before the pathogen can replicate and escape, thereby halting the progression of intracellular infections. T_C cells also support NK lymphocytes to destroy early cancers.

Cytokines secreted by the T_H1 response that stimulates macrophages also stimulate T_C cells and enhance their ability to identify and destroy infected cells and tumors. A summary of how the humoral and cell-mediated immune responses are activated appears in [link].

B plasma cells and T_C cells are collectively called **effector cells** because they are involved in "effecting" (bringing about) the immune response of killing pathogens and infected host cells.

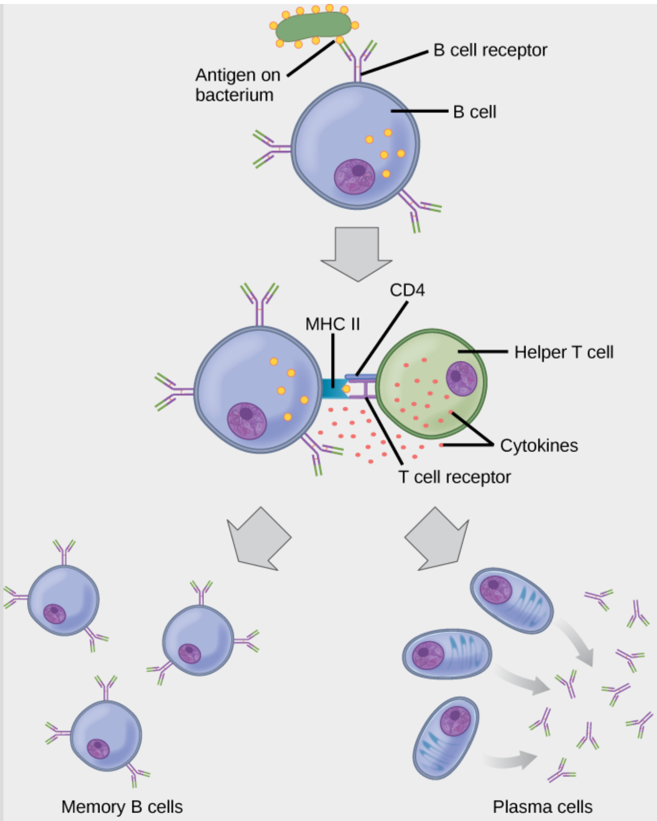


A helper T cell becomes activated by binding to an antigen presented by an APC via the MHCII receptor, causing it to release cytokines. Depending on the cytokines released, this activates either the humoral or the cell-mediated immune response.

Immunological Memory

The adaptive immune system has a memory component that allows for a rapid and large response upon reinvasion of the same pathogen. During the adaptive immune response to a pathogen that has not been encountered before, known as the **primary immune response**, plasma cells secreting antibodies and differentiated T cells increase, then plateau over time. As B and T cells mature into effector cells, a subset of the naïve populations differentiates into B and T memory cells with the same antigen specificities ([link]). A **memory cell** is an antigen-specific B or T lymphocyte that does not differentiate into an effector cell during the primary immune response, but that can immediately become an effector cell on reexposure to the same pathogen. As the infection is cleared and pathogenic stimuli subside, the effectors are no longer needed and they undergo apoptosis. In contrast, the memory cells persist in the circulation.

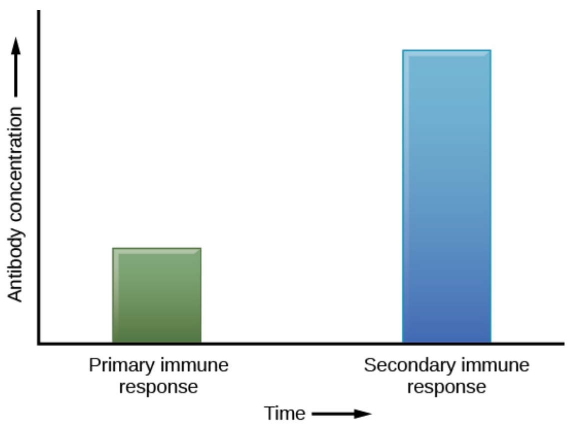
Note:
Art Connection



After initially binding an antigen to the B cell receptor, a B cell internalizes the antigen and presents it on MHC class II. A helper T cell recognizes the MHC class II- antigen complex and activates the B cell. As a result, memory B cells and plasma cells are made.

The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?

If the pathogen is never encountered again during the individual’s lifetime, B and T memory cells will circulate for a few years or even several decades and will gradually die off, having never functioned as effector cells. However, if the host is re-exposed to the same pathogen type, circulating memory cells will immediately differentiate into plasma cells and T_C cells without input from APCs or T_H cells. This is known as the **secondary immune response**. One reason why the adaptive immune response is delayed is because it takes time for naïve B and T cells with the appropriate antigen specificities to be identified, activated, and proliferate. On reinfection, this step is skipped, and the result is a more rapid production of immune defenses. Memory B cells that differentiate into plasma cells output tens to hundreds-fold greater antibody amounts than were secreted during the primary response ([link]). This rapid and dramatic antibody response may stop the infection before it can even become established, and the individual may not realize they had been exposed.



In the primary response to infection, antibodies are secreted first from plasma cells. Upon re-exposure to the same pathogen, memory cells differentiate into antibody-secreting plasma cells that output a greater

amount of antibody for a
longer period of time.

Vaccination is based on the knowledge that exposure to noninfectious antigens, derived from known pathogens, generates a mild primary immune response. The immune response to vaccination may not be perceived by the host as illness but still confers immune memory. When exposed to the corresponding pathogen to which an individual was vaccinated, the reaction is similar to a secondary exposure. Because each reinfection generates more memory cells and increased resistance to the pathogen, some vaccine courses involve one or more booster vaccinations to mimic repeat exposures.

The Lymphatic System

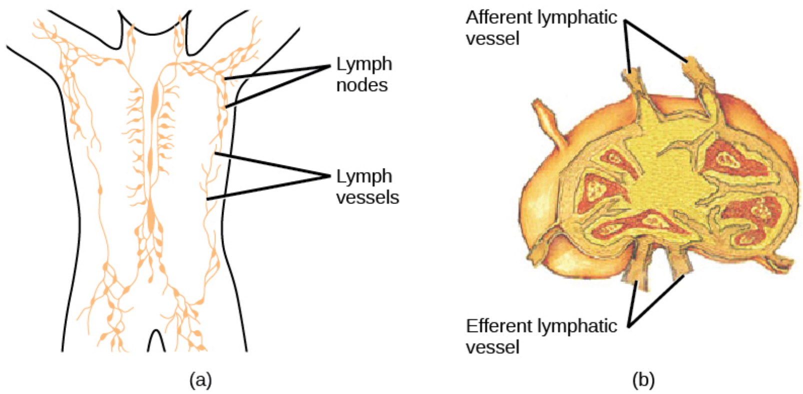
Lymph is the watery fluid that bathes tissues and organs and contains protective white blood cells but does not contain erythrocytes. Lymph moves about the body through the lymphatic system, which is made up of vessels, lymph ducts, lymph glands, and organs, such as tonsils, adenoids, thymus, and spleen.

Although the immune system is characterized by circulating cells throughout the body, the regulation, maturation, and intercommunication of immune factors occur at specific sites. The blood circulates immune cells, proteins, and other factors through the body. Approximately 0.1 percent of all cells in the blood are leukocytes, which include monocytes (the precursor of macrophages) and lymphocytes. Most cells in the blood are red blood cells. Cells of the immune system can travel between the distinct lymphatic and blood circulatory systems, which are separated by interstitial space, by a process called extravasation (passing through to surrounding tissue).

Recall that cells of the immune system originate from stem cells in the bone marrow. B cell maturation occurs in the bone marrow, whereas progenitor

cells migrate from the bone marrow and develop and mature into naïve T cells in the organ called the thymus.

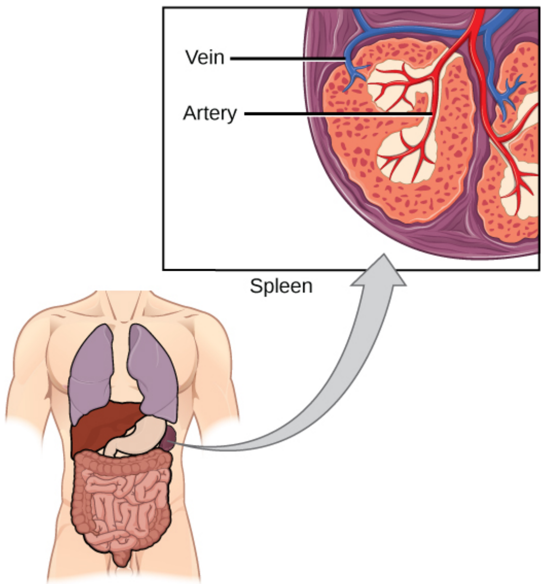
On maturation, T and B lymphocytes circulate to various destinations. Lymph nodes scattered throughout the body house large populations of T and B cells, dendritic cells, and macrophages ([link]). Lymph gathers antigens as it drains from tissues. These antigens then are filtered through lymph nodes before the lymph is returned to circulation. APCs in the lymph nodes capture and process antigens and inform nearby lymphocytes about potential pathogens.



(a) Lymphatic vessels carry a clear fluid called lymph throughout the body. The liquid passes through (b) lymph nodes that filter the lymph that enters the node through afferent vessels and leaves through efferent vessels; lymph nodes are filled with lymphocytes that purge infecting cells. (credit a: modification of work by NIH; credit b: modification of work by NCI, NIH)

The spleen houses B and T cells, macrophages, dendritic cells, and NK cells ([link]). The spleen is the site where APCs that have trapped foreign particles in the blood can communicate with lymphocytes. Antibodies are

synthesized and secreted by activated plasma cells in the spleen, and the spleen filters foreign substances and antibody-complexed pathogens from the blood. Functionally, the spleen is to the blood as lymph nodes are to the lymph.



The spleen functions to immunologically filter the blood and allow for communication between cells corresponding to the innate and adaptive immune responses. (credit: modification of work by NCI, NIH)

Mucosal Immune System

The innate and adaptive immune responses compose the systemic immune system (affecting the whole body), which is distinct from the mucosal

immune system. Mucosa associated lymphoid tissue (MALT) is a crucial component of a functional immune system because mucosal surfaces, such as the nasal passages, are the first tissues onto which inhaled or ingested pathogens are deposited. The mucosal tissue includes the mouth, pharynx, and esophagus, and the gastrointestinal, respiratory, and urogenital tracts.

Mucosal immunity is formed by MALT, which functions independently of the systemic immune system, and which has its own innate and adaptive components. MALT is a collection of lymphatic tissue that combines with epithelial tissue lining the mucosa throughout the body. This tissue functions as the immune barrier and response in areas of the body with direct contact to the external environment. The systemic and mucosal immune systems use many of the same cell types. Foreign particles that make their way to MALT are taken up by absorptive epithelial cells and delivered to APCs located directly below the mucosal tissue. APCs of the mucosal immune system are primarily dendritic cells, with B cells and macrophages having minor roles. Processed antigens displayed on APCs are detected by T cells in the MALT and at the tonsils, adenoids, appendix, or the mesenteric lymph nodes of the intestine. Activated T cells then migrate through the lymphatic system and into the circulatory system to mucosal sites of infection.

Immune Tolerance

The immune system has to be regulated to prevent wasteful, unnecessary responses to harmless substances, and more importantly, so that it does not attack “self.” The acquired ability to prevent an unnecessary or harmful immune response to a detected foreign substance known not to cause disease, or self-antigens, is described as **immune tolerance**. The primary mechanism for developing immune tolerance to self-antigens occurs during the selection for weakly self-binding cells during T and B lymphocyte maturation. There are populations of T cells that suppress the immune response to self-antigens and that suppress the immune response after the infection has cleared to minimize host cell damage induced by inflammation and cell lysis. Immune tolerance is especially well developed in the mucosa of the upper digestive system because of the tremendous number of foreign substances (such as food proteins) that APCs of the oral

cavity, pharynx, and gastrointestinal mucosa encounter. Immune tolerance is brought about by specialized APCs in the liver, lymph nodes, small intestine, and lung that present harmless antigens to a diverse population of regulatory T (T_{reg}) cells, specialized lymphocytes that suppress local inflammation and inhibit the secretion of stimulatory immune factors. The combined result of T_{reg} cells is to prevent immunologic activation and inflammation in undesired tissue compartments and to allow the immune system to focus on pathogens instead.

Section Summary

The adaptive immune response is a slower-acting, longer-lasting, and more specific response than the innate response. However, the adaptive response requires information from the innate immune system to function. APCs display antigens on MHC molecules to naïve T cells. T cells with cell-surface receptors that bind a specific antigen will bind to that APC. In response, the T cells differentiate and proliferate, becoming T_H cells or T_C cells. T_H cells stimulate B cells that have engulfed and presented pathogen-derived antigens. B cells differentiate into plasma cells that secrete antibodies, whereas T_C cells destroy infected or cancerous cells. Memory cells are produced by activated and proliferating B and T cells and persist after a primary exposure to a pathogen. If re-exposure occurs, memory cells differentiate into effector cells without input from the innate immune system. The mucosal immune system is largely independent of the systemic immune system but functions in parallel to protect the extensive mucosal surfaces of the body. Immune tolerance is brought about by T_{reg} cells to limit reactions to harmless antigens and the body's own molecules.

Art Connections

Exercise:

Problem:

[link] The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?

Solution:

[link] If the blood of the mother and fetus mixes, memory cells that recognize the Rh antigen of the fetus can form in the mother late in the first pregnancy. During subsequent pregnancies, these memory cells launch an immune attack on the fetal blood cells of an Rh-positive fetus. Injection of anti-Rh antibody during the first pregnancy prevents the immune response from occurring.

Review Questions

Exercise:

Problem: The humoral immune response depends on which cells?

a. T_C cells
b. B cells
c. B and T_H cells
d. T_C and T_H cells

Solution:

C

Exercise:

Problem:

The fact that the body does not normally mount an immune response to the molecules in food is an example of _______.

a. secondary immune response
b. immunological memory
c. immune tolerance
d. passive immunity

Solution:

C

Exercise:

Problem:

Foreign particles circulating in the blood are filtered by the ____________.

a. spleen
b. lymph nodes
c. MALT
d. lymph

Solution:

A

Free Response

Exercise:

Problem:

How do B and T cells differ with respect to antigens that they bind?

Solution:

T cells bind antigens that have been digested and embedded in MHC molecules by APCs. In contrast, B cells function as APCs to bind intact, unprocessed antigens.

Exercise:

Problem:

Why is the immune response after reinfection much faster than the adaptive immune response after the initial infection?

Solution:

Upon reinfection, the memory cells will immediately differentiate into plasma cells and CTLs without input from APCs or T_H cells. In contrast, the adaptive immune response to the initial infection requires time for naïve B and T cells with the appropriate antigen specificities to be identified and activated.

Glossary

active immunity
: an immunity that occurs as a result of the activity of the body’s own cells rather than from antibodies acquired from an external source

adaptive immunity
: a specific immune response that occurs after exposure to an antigen either from a pathogen or a vaccination

antibody
: a protein that is produced by plasma cells after stimulation by an antigen; also known as an immunoglobulin

antigen

a macromolecule that reacts with cells of the immune system and which may or may not have a stimulatory effect

antigen-presenting cell (APC)
: an immune cell that detects, engulfs, and informs the adaptive immune response about an infection by presenting the processed antigen on its cell surface

B cell
: a lymphocyte that matures in the bone marrow

cell-mediated immune response
: an adaptive immune response that is controlled by T cells

cytotoxic T lymphocyte (T_C)
: an adaptive immune cell that directly kills infected cells via enzymes, and that releases cytokines to enhance the immune response

dendritic cell
: an immune cell that processes antigen material and presents it on the surface of its cell in MHC class II molecules and induces an immune response in other cells

effector cell
: a lymphocyte that has differentiated, such as a B cell, plasma cell, or cytotoxic T cell

helper T lymphocyte (T_H)
: a cell of the adaptive immune system that binds APCs via MHC class II molecules and stimulates B cells or secretes cytokines to initiate the immune response

humoral immune response
: the adaptive immune response that is controlled by activated B cells and antibodies

immune tolerance

an acquired ability to prevent an unnecessary or harmful immune response to a detected foreign body known not to cause disease

lymph
: the watery fluid present in the lymphatic circulatory system that bathes tissues and organs with protective white blood cells and does not contain erythrocytes

memory cell
: an antigen-specific B or T lymphocyte that does not differentiate into an effector cell during the primary immune response but that can immediately become an effector cell on reexposure to the same pathogen

major histocompatibility class (MHC) II molecule
: a protein found on the surface of antigen-presenting cells that signals to immune cells whether the cell is normal or is infected or cancerous; it provides the appropriate template into which antigens can be loaded for recognition by lymphocytes

passive immunity
: an immunity that does not result from the activity of the body’s own immune cells but by transfer of antibodies from one individual to another

primary immune response
: the response of the adaptive immune system to the first exposure to an antigen

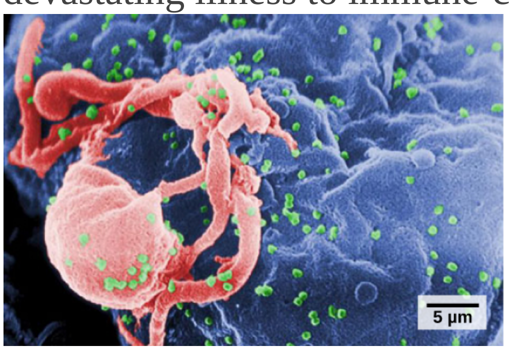
secondary immune response
: the response of the adaptive immune system to a second or later exposure to an antigen mediated by memory cells

T cell
: a lymphocyte that matures in the thymus gland

Disruptions in the Immune System
By the end of this section, you will be able to:

- Describe hypersensitivity
- Define autoimmunity

A functioning immune system is essential for survival, but even the sophisticated cellular and molecular defenses of the mammalian immune response can be defeated by pathogens at virtually every step. In the competition between immune protection and pathogen evasion, pathogens have the advantage of more rapid evolution because of their shorter generation time, large population sizes and often higher mutation rates. Thus pathogens have evolved a diverse array of immune escape mechanisms. For instance, *Streptococcus pneumoniae* (the bacterium that causes pneumonia and meningitis) surrounds itself with a capsule that inhibits phagocytes from engulfing it and displaying antigens to the adaptive immune system. *Staphylococcus aureus* (the bacterium that can cause skin infections, abscesses, and meningitis) synthesizes a toxin called leukocidin that kills phagocytes after they engulf the bacterium. Other pathogens can also hinder the adaptive immune system. HIV infects T_H cells using their CD4 surface molecules, gradually depleting the number of T_H cells in the body ([link]); this inhibits the adaptive immune system's capacity to generate sufficient responses to infection or tumors. As a result, HIV-infected individuals often suffer from infections that would not cause illness in people with healthy immune systems but which can cause devastating illness to immune-compromised individuals.



HIV (green) is shown budding from a lymphocyte cell (red) in

culture. (credit: modification of work by C. Goldsmith, CDC; scale-bar data from Matt Russell)

Inappropriate responses of immune cells and molecules themselves can also disrupt the proper functioning of the entire system, leading to host-cell damage that can become fatal.

Immunodeficiency

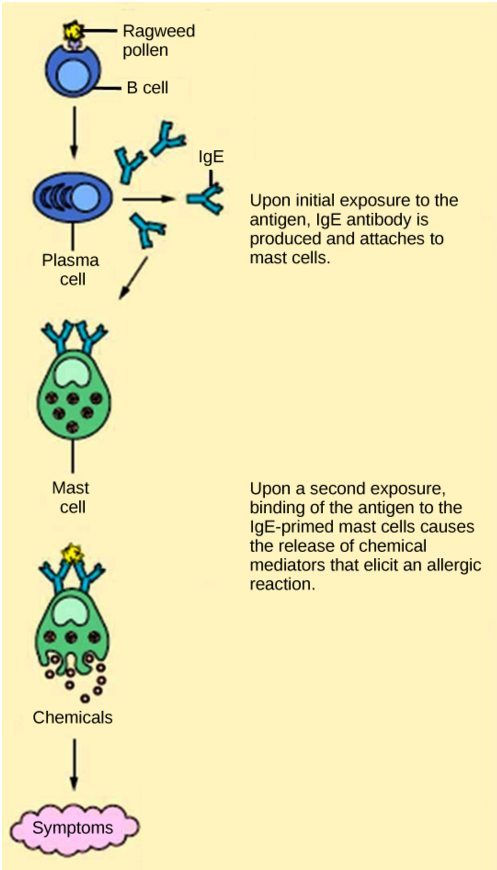
Immunodeficiency is a failure, insufficiency, or delay in the response of the immune system, which may be acquired or inherited. Immunodeficiency can allow pathogens or tumor cells to gain a foothold and replicate or proliferate to high enough levels so that the immune system becomes overwhelmed. Immunodeficiency can be acquired as a result of infection with certain pathogens that attack the cells of the immune system itself (such as HIV), chemical exposure (including certain medical treatments such as chemotherapy), malnutrition, or extreme stress. For instance, radiation exposure can destroy populations of lymphocytes and elevate an individual's susceptibility to infections and cancer. Rarely, primary immunodeficiencies that are present from birth may also occur. For example, severe combined immunodeficiency disease (SCID) is a condition in which children are born without functioning B or T cells.

Hypersensitivities

A maladaptive immune response toward harmless foreign substances or self-antigens that occur after tissue sensitization is termed a **hypersensitivity**. Types of hypersensitivities include immediate, delayed, and autoimmune. A large proportion of the human population is affected by one or more types of hypersensitivity.

Allergies

The immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a usually harmless antigen is called an **allergy**. In the United States, 20 percent of the population exhibits symptoms of allergy or asthma, whereas 55 percent test positive against one or more allergens. On initial exposure to a potential allergen, an allergic individual synthesizes antibodies through the typical process of APCs presenting processed antigen to T_H cells that stimulate B cells to produce the antibodies. The antibody molecules interact with mast cells embedded in connective tissues. This process primes, or sensitizes, the tissue. On subsequent exposure to the same allergen, antibody molecules on mast cells bind the antigen and stimulate the mast cell to release histamine and other inflammatory chemicals; these chemical mediators then recruit eosinophils (a type of white blood cell), which also appear to be adapted to responding to parasitic worms ([link]). Eosinophils release factors that enhance the inflammatory response and the secretions of mast cells. The effects of an allergic reaction range from mild symptoms like sneezing and itchy, watery eyes to more severe or even life-threatening reactions involving intensely itchy welts or hives, airway constriction with severe respiratory distress, and plummeting blood pressure caused by dilating blood vessels and fluid loss from the circulatory system. This extreme reaction, typically in response to an allergen introduced to the circulatory system, is known as anaphylactic shock. Antihistamines are an insufficient counter to anaphylactic shock and if not treated with epinephrine to counter the blood pressure and breathing effects, this condition can be fatal.



On first exposure to an allergen, an antibody is synthesized by plasma cells in response to a harmless antigen. The antibodies bind to mast cells, and on secondary exposure, the mast cells release histamines and other modulators that cause the symptoms of allergy. (credit: modification of work by NIH)

Delayed hypersensitivity is a cell-mediated immune response that takes approximately one to two days after secondary exposure for a maximal reaction. This type of hypersensitivity involves the T_H1 cytokine-mediated inflammatory response and may cause local tissue lesions or contact dermatitis (rash or skin irritation). Delayed hypersensitivity occurs in some individuals in response to contact with certain types of jewelry or cosmetics. Delayed hypersensitivity facilitates the immune response to poison ivy and is also the reason why the skin test for tuberculosis results in a small region of inflammation on individuals who were previously exposed to *Mycobacterium tuberculosis*, the organism that causes tuberculosis.

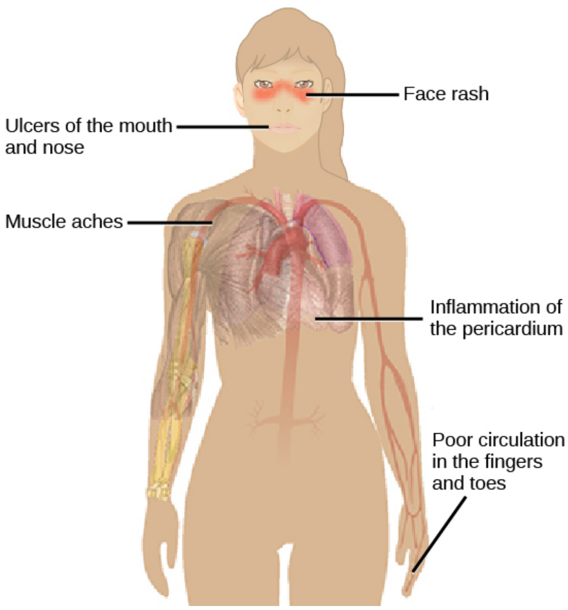
Note:
Concept in Action

Try your hand at diagnosing an allergic reaction by selecting one of the interactive case studies at the World Allergy Organization website.

Autoimmunity

Autoimmunity is a type of hypersensitivity to self-antigens that affects approximately five percent of the population. Most types of autoimmunity involve the humoral immune response. An antibody that inappropriately marks self-components as foreign is termed an **autoantibody**. In patients with myasthenia gravis, an autoimmune disease, muscle-cell receptors that induce contraction in response to acetylcholine are targeted by antibodies. The result is muscle weakness that may include marked difficultly with fine or gross motor functions. In systemic lupus erythematosus, a diffuse autoantibody response to the individual’s own DNA and proteins results in

various systemic diseases ([link]). Systemic lupus erythematosus may affect the heart, joints, lungs, skin, kidneys, central nervous system, or other tissues, causing tissue damage through antibody binding, complement recruitment, lysis, and inflammation.



Systemic lupus erythematosus is characterized by autoimmunity to the individual's own DNA and/or proteins, which leads to varied dysfunction of the organs. (credit: modification of work by Mikael Häggström)

Autoimmunity can develop with time and its causes may be rooted in molecular mimicry, a situation in which one molecule is similar enough in shape to another molecule that it binds the same immune receptors.

Antibodies and T-cell receptors may bind self-antigens that are structurally similar to pathogen antigens. As an example, infection with *Streptococcus pyogenes* (the bacterium that causes strep throat) may generate antibodies or T cells that react with heart muscle, which has a similar structure to the surface of *S. pyogenes*. These antibodies can damage heart muscle with autoimmune attacks, leading to rheumatic fever. Insulin-dependent (Type 1) diabetes mellitus arises from a destructive inflammatory T_H1 response against insulin-producing cells of the pancreas. Patients with this autoimmunity must be treated with regular insulin injections.

Section Summary

Immune disruptions may involve insufficient immune responses or inappropriate immune responses. Immunodeficiency increases an individual's susceptibility to infections and cancers. Hypersensitivities are misdirected responses either to harmless foreign particles, as in the case of allergies, or to the individual’s own tissues, as in the case of autoimmunity. Reactions to self-components may be the result of molecular mimicry.

Review Questions

Exercise:

Problem: Allergy to pollen is classified as ________.

a. an autoimmune reaction
b. immunodeficiency
c. delayed hypersensitivity
d. immediate hypersensitivity

Solution:

D

Exercise:

Problem:A potential cause of acquired autoimmunity is ________.

a. tissue hypersensitivity
b. molecular mimicry
c. histamine release
d. radiation exposure

Solution:

B

Exercise:

Problem: Autoantibodies are probably involved in ________.

a. reactions to poison ivy
b. pollen allergies
c. systemic lupus erythematosus
d. HIV/AIDS

Solution:

C

Free Response

Exercise:

Problem:

Some photographers develop a sensitivity to certain film developing chemicals leading to severe rashes on their hands such that they are unable to work with them. Explain what is probably happening.

Solution:

This is probably a delayed sensitivity reaction to one or more chemicals in the developer. An initial exposure would have sensitized the individual to the chemical and then subsequent exposures will induce a delayed inflammation reaction a day or two after exposure.

Glossary

allergy
: an immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a harmless antigen

autoantibody
: an antibody that incorrectly marks “self” components as foreign and stimulates the immune response

autoimmunity
: a type of hypersensitivity to self-antigens

hypersensitivity
: a spectrum of inappropriate immune responses toward harmless foreign particles or self-antigens; occurs after tissue sensitization and includes immediate-type (allergy), delayed-type, and autoimmunity

immunodeficiency
: a failure, insufficiency, or delay at any level of the immune system, which may be acquired or inherited

Types of Skeletal Systems
By the end of this section, you will be able to:

- Discuss the different types of skeletal systems
- Explain the role of the human skeletal system
- Compare and contrast different skeletal systems

A skeletal system is necessary to support the body, protect internal organs, and allow for the movement of an organism. There are three different skeleton designs that fulfill these functions: hydrostatic skeleton, exoskeleton, and endoskeleton.

Hydrostatic Skeleton

A **hydrostatic skeleton** is a skeleton formed by a fluid-filled compartment within the body, called the coelom. The organs of the coelom are supported by the aqueous fluid, which also resists external compression. This compartment is under hydrostatic pressure because of the fluid and supports the other organs of the organism. This type of skeletal system is found in soft-bodied animals such as sea anemones, earthworms, Cnidaria, and other invertebrates ([link]).

The skeleton of the red-knobbed sea star (*Protoreaster linckii*) is an example of a hydrostatic skeleton.

(credit: “Amada44”/Wikimedia Commons)

Movement in a hydrostatic skeleton is provided by muscles that surround the coelom. The muscles in a hydrostatic skeleton contract to change the shape of the coelom; the pressure of the fluid in the coelom produces movement. For example, earthworms move by waves of muscular contractions of the skeletal muscle of the body wall hydrostatic skeleton, called peristalsis, which alternately shorten and lengthen the body. Lengthening the body extends the anterior end of the organism. Most organisms have a mechanism to fix themselves in the substrate. Shortening the muscles then draws the posterior portion of the body forward. Although a hydrostatic skeleton is well-suited to invertebrate organisms such as earthworms and some aquatic organisms, it is not an efficient skeleton for terrestrial animals.

Exoskeleton

An **exoskeleton** is an external skeleton that consists of a hard encasement on the surface of an organism. For example, the shells of crabs and insects are exoskeletons ([link]). This skeleton type provides defence against predators, supports the body, and allows for movement through the contraction of attached muscles. As with vertebrates, muscles must cross a joint inside the exoskeleton. Shortening of the muscle changes the relationship of the two segments of the exoskeleton. Arthropods such as crabs and lobsters have exoskeletons that consist of 30–50 percent chitin, a polysaccharide derivative of glucose that is a strong but flexible material. Chitin is secreted by the epidermal cells. The exoskeleton is further strengthened by the addition of calcium carbonate in organisms such as the lobster. Because the exoskeleton is acellular, arthropods must periodically shed their exoskeletons because the exoskeleton does not grow as the organism grows.

Muscles attached to the exoskeleton of the Halloween crab (*Gecarcinus quadratus*) allow it to move.

Endoskeleton

An **endoskeleton** is a skeleton that consists of hard, mineralized structures located within the soft tissue of organisms. An example of a primitive endoskeletal structure is the spicules of sponges. The bones of vertebrates are composed of tissues, whereas sponges have no true tissues ([link]). Endoskeletons provide support for the body, protect internal organs, and allow for movement through contraction of muscles attached to the skeleton.

The skeletons of humans and horses are examples of endoskeletons. (credit: Ross Murphy)

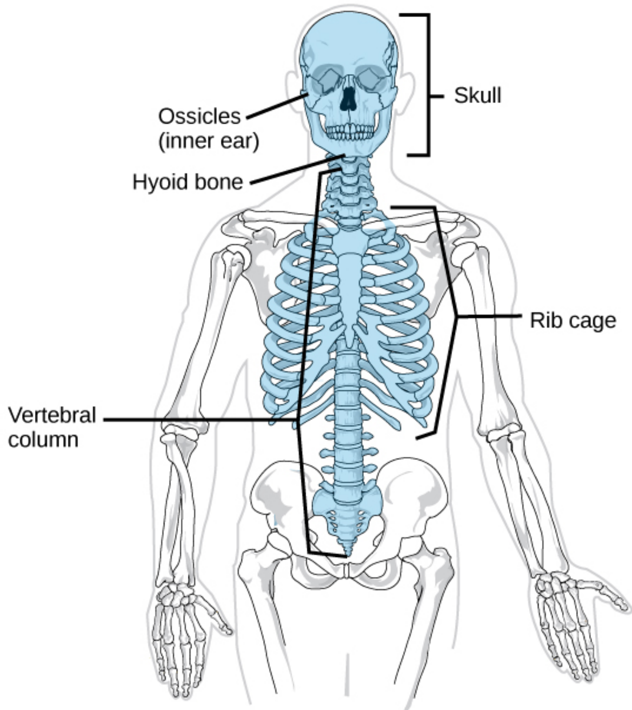
The human skeleton is an endoskeleton that consists of 206 bones in the adult. It has five main functions: providing support to the body, storing minerals and lipids, producing blood cells, protecting internal organs, and allowing for movement. The skeletal system in vertebrates is divided into the axial skeleton (which consists of the skull, vertebral column, and rib cage), and the appendicular skeleton (which consists of the shoulders, limb bones, the pectoral girdle, and the pelvic girdle).

Note:
Link to Learning

Visit the interactive body site to build a virtual skeleton: select "skeleton" and click through the activity to place each bone.

Human Axial Skeleton

The **axial skeleton** forms the central axis of the body and includes the bones of the skull, ossicles of the middle ear, hyoid bone of the throat, vertebral column, and the thoracic cage (ribcage) ([link]). The function of the axial skeleton is to provide support and protection for the brain, the spinal cord, and the organs in the ventral body cavity. It provides a surface for the attachment of muscles that move the head, neck, and trunk, performs respiratory movements, and stabilizes parts of the appendicular skeleton.

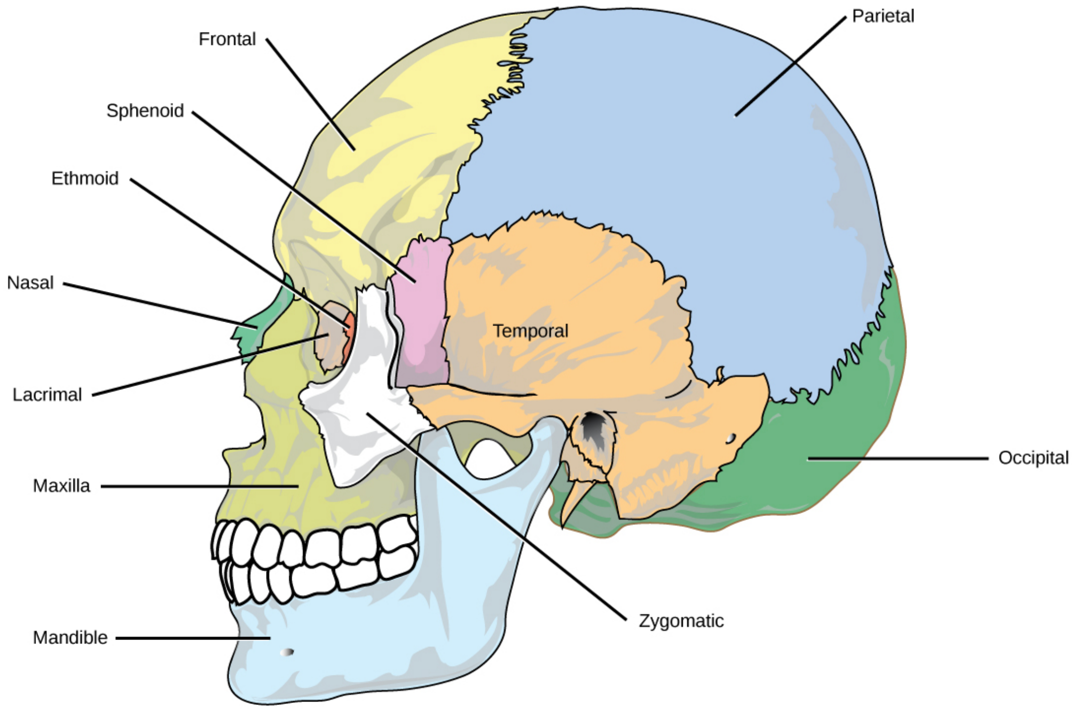


The axial skeleton consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and rib cage. (credit: modification of work by Mariana Ruiz Villareal)

The Skull

The bones of the **skull** support the structures of the face and protect the brain. The skull consists of 22 bones, which are divided into two categories: cranial bones and facial bones. The **cranial bones** are eight bones that form the cranial cavity, which encloses the brain and serves as an attachment site for the muscles of the head and neck. The eight cranial bones are the frontal bone, two parietal bones, two temporal bones, occipital bone, sphenoid bone, and the ethmoid bone. Although the bones developed separately in

the embryo and fetus, in the adult, they are tightly fused with connective tissue and adjoining bones do not move ([link]).

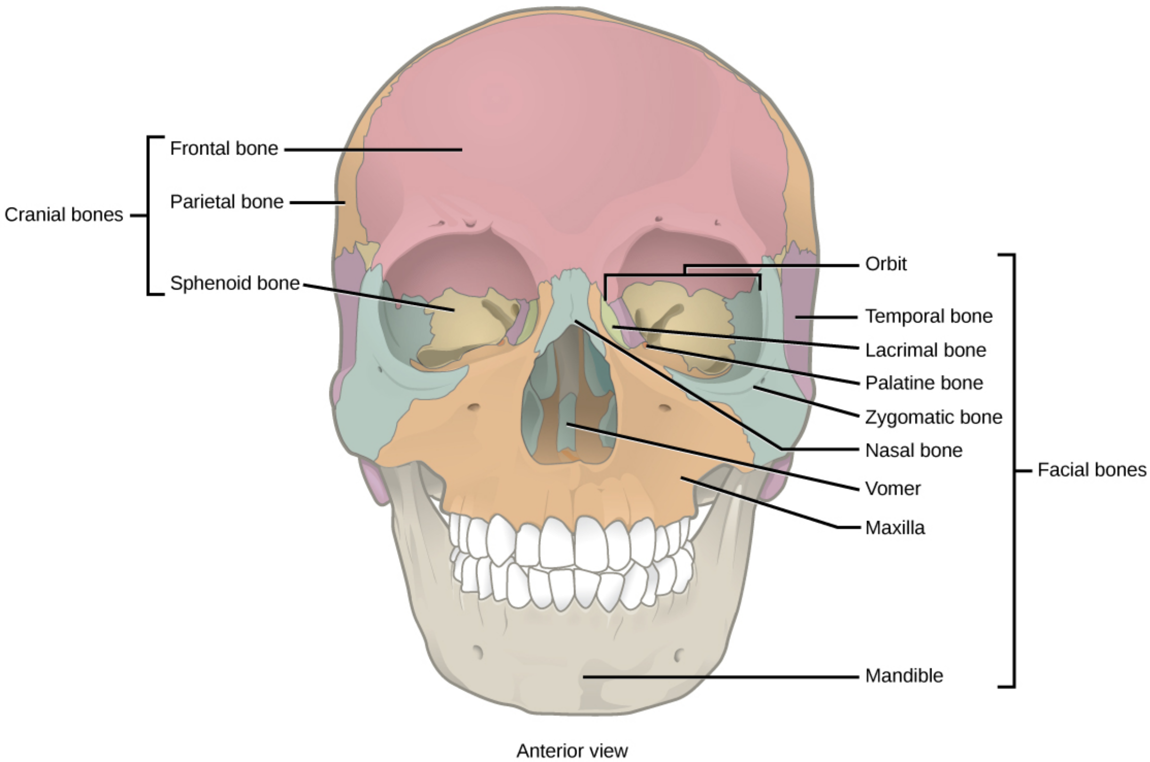


The bones of the skull support the structures of the face and protect the brain. (credit: modification of work by Mariana Ruiz Villareal)

The **auditory ossicles** of the middle ear transmit sounds from the air as vibrations to the fluid-filled cochlea. The auditory ossicles consist of three bones each: the malleus, incus, and stapes. These are the smallest bones in the body and are unique to mammals.

Fourteen **facial bones** form the face, provide cavities for the sense organs (eyes, mouth, and nose), protect the entrances to the digestive and respiratory tracts, and serve as attachment points for facial muscles. The 14 facial bones are the nasal bones, the maxillary bones, zygomatic bones, palatine, vomer, lacrimal bones, the inferior nasal conchae, and the

mandible. All of these bones occur in pairs except for the mandible and the vomer ([link]).

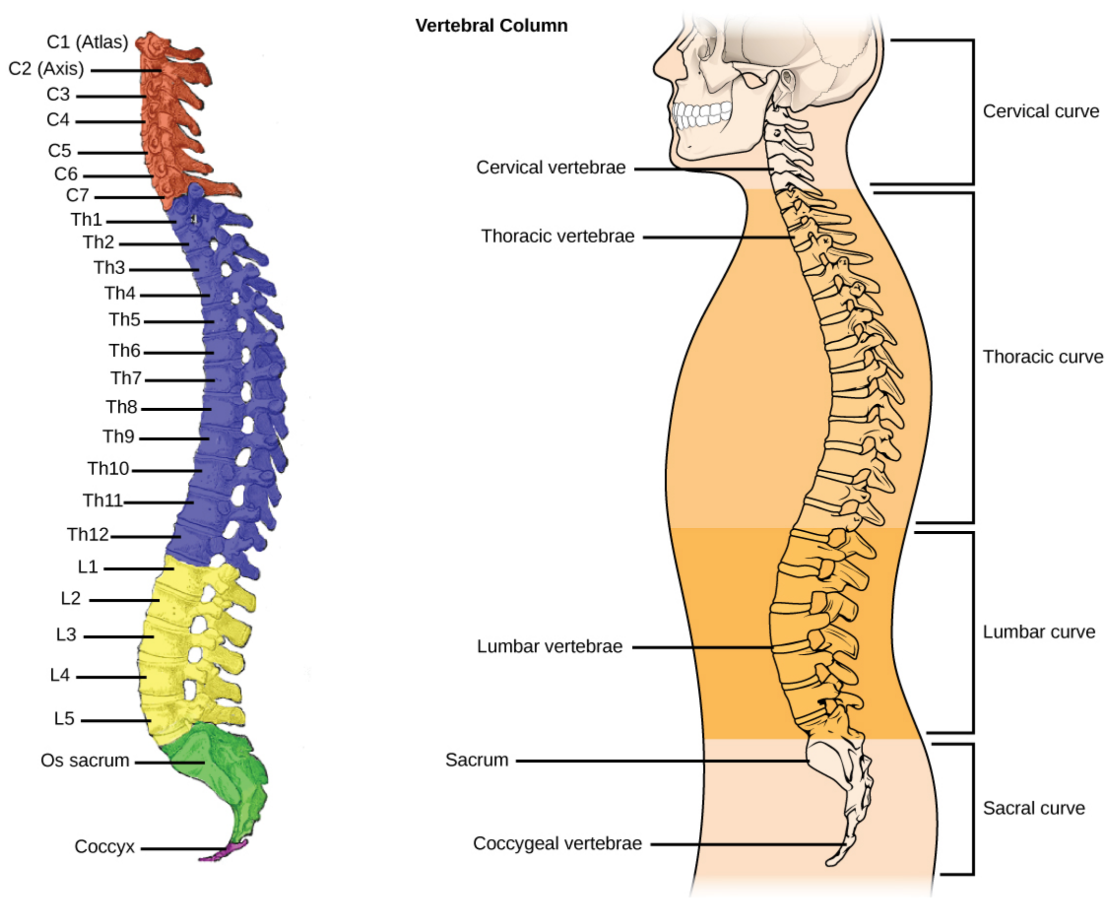


The cranial bones, including the frontal, parietal, and sphenoid bones, cover the top of the head. The facial bones of the skull form the face and provide cavities for the eyes, nose, and mouth.

Although it is not found in the skull, the hyoid bone is considered a component of the axial skeleton. The **hyoid bone** lies below the mandible in the front of the neck. It acts as a movable base for the tongue and is connected to muscles of the jaw, larynx, and tongue. The mandible articulates with the base of the skull. The mandible controls the opening to the airway and gut. In animals with teeth, the mandible brings the surfaces of the teeth in contact with the maxillary teeth.

The Vertebral Column

The **vertebral column**, or spinal column, surrounds and protects the spinal cord, supports the head, and acts as an attachment point for the ribs and muscles of the back and neck. The adult vertebral column comprises 26 bones: the 24 vertebrae, the sacrum, and the coccyx bones. In the adult, the sacrum is typically composed of five vertebrae that fuse into one. The coccyx is typically 3–4 vertebrae that fuse into one. Around the age of 70, the sacrum and the coccyx may fuse together. We begin life with approximately 33 vertebrae, but as we grow, several vertebrae fuse together. The adult vertebrae are further divided into the 7 cervical vertebrae, 12 thoracic vertebrae, and 5 lumbar vertebrae ([link]).



(a) The vertebral column consists of seven cervical vertebrae (C1–7) twelve thoracic vertebrae (Th1–12), five lumbar

vertebrae (L1–5), the os sacrum, and the coccyx. (b) Spinal curves increase the strength and flexibility of the spine. (credit a: modification of work by Uwe Gille based on original work by Gray's Anatomy; credit b: modification of work by NCI, NIH)

Each vertebral body has a large hole in the center through which the nerves of the spinal cord pass. There is also a notch on each side through which the spinal nerves, which serve the body at that level, can exit from the spinal cord. The vertebral column is approximately 71 cm (28 inches) in adult male humans and is curved, which can be seen from a side view. The names of the spinal curves correspond to the region of the spine in which they occur. The thoracic and sacral curves are concave (curve inwards relative to the front of the body) and the cervical and lumbar curves are convex (curve outwards relative to the front of the body). The arched curvature of the vertebral column increases its strength and flexibility, allowing it to absorb shocks like a spring ([link]).

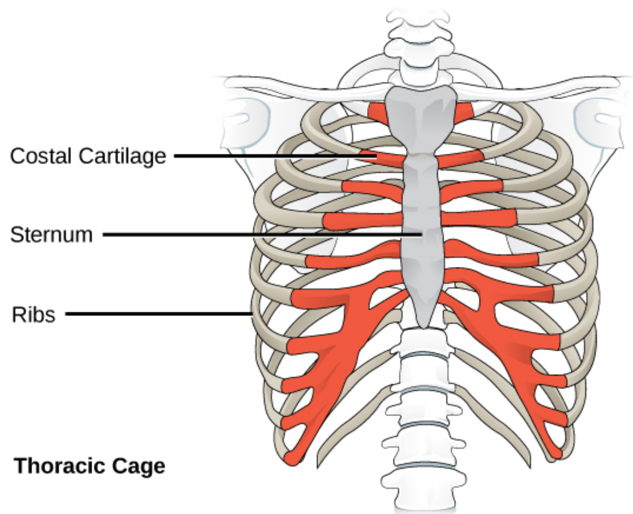
Intervertebral discs composed of fibrous cartilage lie between adjacent vertebral bodies from the second cervical vertebra to the sacrum. Each disc is part of a joint that allows for some movement of the spine and acts as a cushion to absorb shocks from movements such as walking and running. Intervertebral discs also act as ligaments to bind vertebrae together. The inner part of discs, the nucleus pulposus, hardens as people age and becomes less elastic. This loss of elasticity diminishes its ability to absorb shocks.

The Thoracic Cage

The **thoracic cage**, also known as the ribcage, is the skeleton of the chest, and consists of the ribs, sternum, thoracic vertebrae, and costal cartilages ([link]). The thoracic cage encloses and protects the organs of the thoracic cavity, including the heart and lungs. It also provides support for the shoulder girdles and upper limbs, and serves as the attachment point for the

diaphragm, muscles of the back, chest, neck, and shoulders. Changes in the volume of the thorax enable breathing.

The **sternum**, or breastbone, is a long, flat bone located at the anterior of the chest. It is formed from three bones that fuse in the adult. The **ribs** are 12 pairs of long, curved bones that attach to the thoracic vertebrae and curve toward the front of the body, forming the ribcage. Costal cartilages connect the anterior ends of the ribs to the sternum, with the exception of rib pairs 11 and 12, which are free-floating ribs.

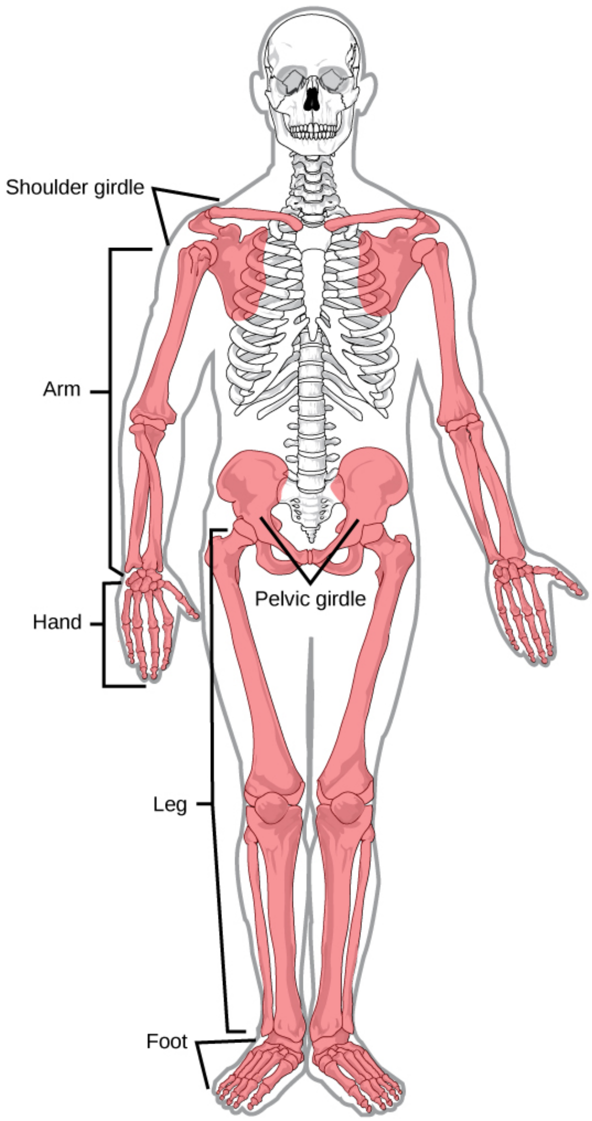


The thoracic cage, or rib cage, protects the heart and the lungs. (credit: modification of work by NCI, NIH)

Human Appendicular Skeleton

The **appendicular skeleton** is composed of the bones of the upper limbs (which function to grasp and manipulate objects) and the lower limbs (which permit locomotion). It also includes the pectoral girdle, or shoulder

girdle, that attaches the upper limbs to the body, and the pelvic girdle that attaches the lower limbs to the body ([link]).

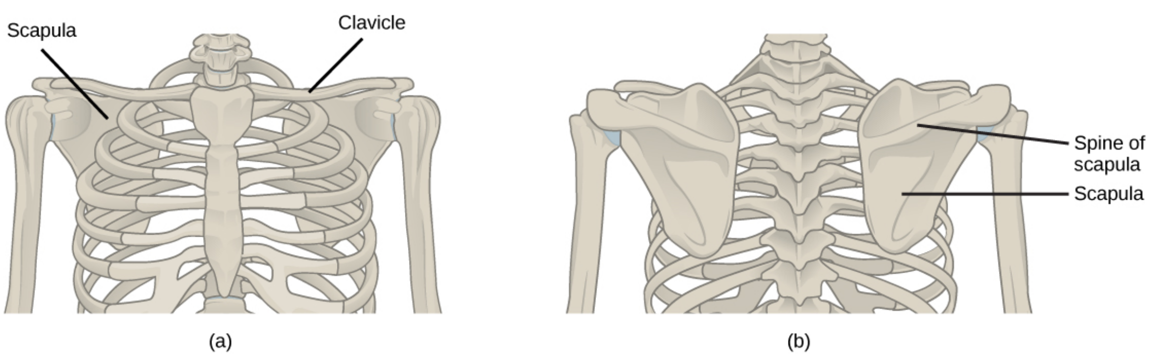


The appendicular skeleton is composed of the bones of the pectoral limbs (arm, forearm, hand), the pelvic limbs (thigh, leg, foot), the pectoral girdle, and the pelvic girdle. (credit:

modification of work by
Mariana Ruiz Villareal)

The Pectoral Girdle

The **pectoral girdle** bones provide the points of attachment of the upper limbs to the axial skeleton. The human pectoral girdle consists of the clavicle (or collarbone) in the anterior, and the scapula (or shoulder blades) in the posterior ([link]).



(a) The pectoral girdle in primates consists of the clavicles and scapulae. (b) The posterior view reveals the spine of the scapula to which muscle attaches.

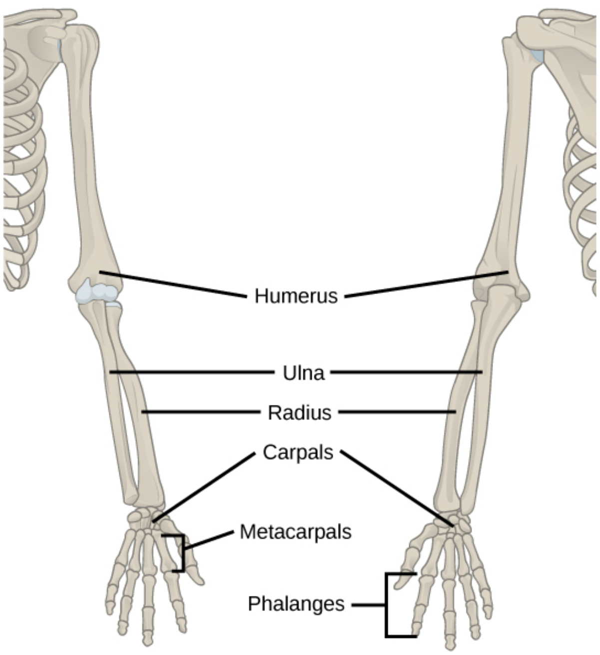
The **clavicles** are S-shaped bones that position the arms on the body. The clavicles lie horizontally across the front of the thorax (chest) just above the first rib. These bones are fairly fragile and are susceptible to fractures. For example, a fall with the arms outstretched causes the force to be transmitted to the clavicles, which can break if the force is excessive. The clavicle articulates with the sternum and the scapula.

The **scapulae** are flat, triangular bones that are located at the back of the pectoral girdle. They support the muscles crossing the shoulder joint. A

ridge, called the spine, runs across the back of the scapula and can easily be felt through the skin ([link]). The spine of the scapula is a good example of a bony protrusion that facilitates a broad area of attachment for muscles to bone.

The Upper Limb

The upper limb contains 30 bones in three regions: the arm (shoulder to elbow), the forearm (ulna and radius), and the wrist and hand ([link]).



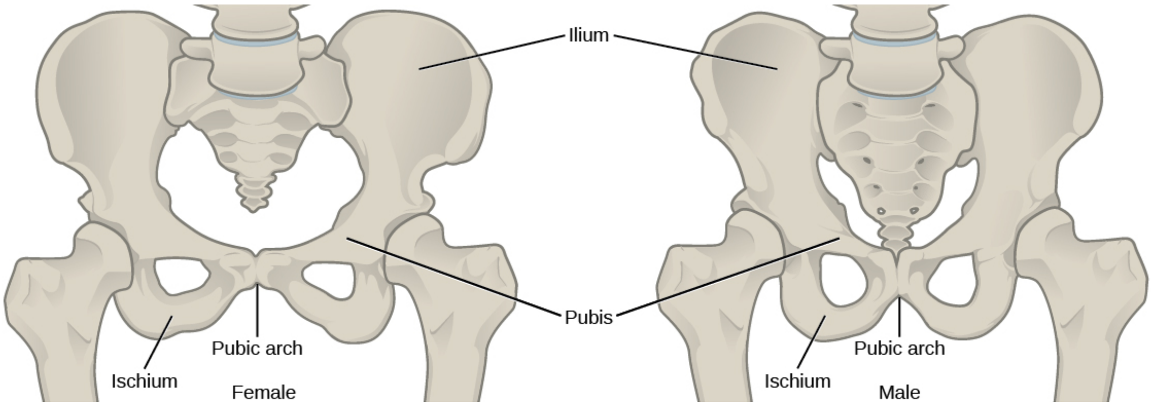
The upper limb consists of the humerus of the upper arm, the radius and ulna of the forearm, eight bones of the carpus, five bones of the metacarpus, and 14 bones of the phalanges.

An **articulation** is any place at which two bones are joined. The **humerus** is the largest and longest bone of the upper limb and the only bone of the arm. It articulates with the scapula at the shoulder and with the forearm at the elbow. The **forearm** extends from the elbow to the wrist and consists of two bones: the ulna and the radius. The **radius** is located along the lateral (thumb) side of the forearm and articulates with the humerus at the elbow. The **ulna** is located on the medial aspect (pinky-finger side) of the forearm. It is longer than the radius. The ulna articulates with the humerus at the elbow. The radius and ulna also articulate with the carpal bones and with each other, which in vertebrates enables a variable degree of rotation of the carpus with respect to the long axis of the limb. The hand includes the eight bones of the **carpus** (wrist), the five bones of the **metacarpus** (palm), and the 14 bones of the **phalanges** (digits). Each digit consists of three phalanges, except for the thumb, when present, which has only two.

The Pelvic Girdle

The **pelvic girdle** attaches to the lower limbs of the axial skeleton. Because it is responsible for bearing the weight of the body and for locomotion, the pelvic girdle is securely attached to the axial skeleton by strong ligaments. It also has deep sockets with robust ligaments to securely attach the femur to the body. The pelvic girdle is further strengthened by two large hip bones. In adults, the hip bones, or **coxal bones**, are formed by the fusion of three pairs of bones: the ilium, ischium, and pubis. The pelvis joins together in the anterior of the body at a joint called the pubic symphysis and with the bones of the sacrum at the posterior of the body.

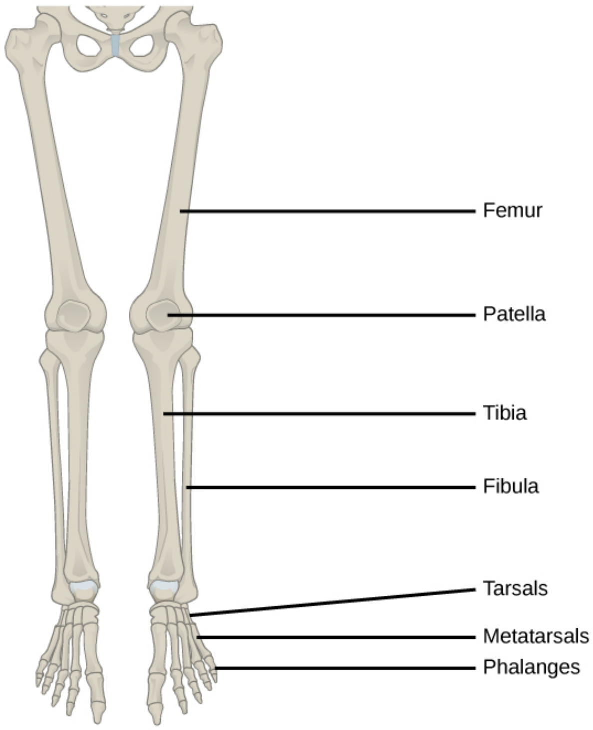
The female pelvis is slightly different from the male pelvis. Over generations of evolution, females with a wider pubic angle and larger diameter pelvic canal reproduced more successfully. Therefore, their offspring also had pelvic anatomy that enabled successful childbirth ([link]).



To adapt to reproductive fitness, the (a) female pelvis is lighter, wider, shallower, and has a broader angle between the pubic bones than (b) the male pelvis.

The Lower Limb

The **lower limb** consists of the thigh, the leg, and the foot. The bones of the lower limb are the femur (thigh bone), patella (kneecap), tibia and fibula (bones of the leg), tarsals (bones of the ankle), and metatarsals and phalanges (bones of the foot) ([link]). The bones of the lower limbs are thicker and stronger than the bones of the upper limbs because of the need to support the entire weight of the body and the resulting forces from locomotion. In addition to evolutionary fitness, the bones of an individual will respond to forces exerted upon them.

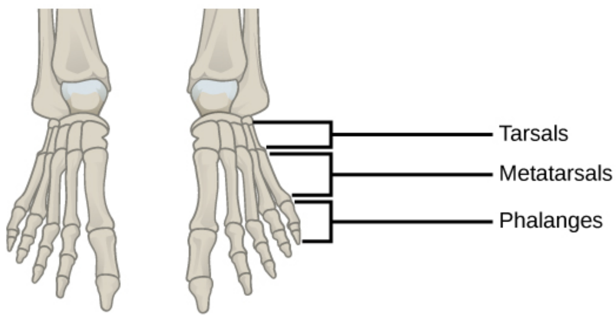


The lower limb consists of the thigh (femur), kneecap (patella), leg (tibia and fibula), ankle (tarsals), and foot (metatarsals and phalanges) bones.

The **femur**, or thighbone, is the longest, heaviest, and strongest bone in the body. The femur and pelvis form the hip joint at the proximal end. At the distal end, the femur, tibia, and patella form the knee joint. The **patella**, or kneecap, is a triangular bone that lies anterior to the knee joint. The patella is embedded in the tendon of the femoral extensors (quadriceps). It improves knee extension by reducing friction. The **tibia**, or shinbone, is a large bone of the leg that is located directly below the knee. The tibia articulates with the femur at its proximal end, with the fibula and the tarsal bones at its distal end. It is the second largest bone in the human body and is responsible for transmitting the weight of the body from the femur to the foot. The **fibula**, or calf bone, parallels and articulates with the tibia. It does

not articulate with the femur and does not bear weight. The fibula acts as a site for muscle attachment and forms the lateral part of the ankle joint.

The **tarsals** are the seven bones of the ankle. The ankle transmits the weight of the body from the tibia and the fibula to the foot. The **metatarsals** are the five bones of the foot. The phalanges are the 14 bones of the toes. Each toe consists of three phalanges, except for the big toe that has only two ([link]). Variations exist in other species; for example, the horse's metacarpals and metatarsals are oriented vertically and do not make contact with the substrate.



This drawing shows the bones of the human foot and ankle, including the metatarsals and the phalanges.

Note:
Evolution Connection
Evolution of Body Design for Locomotion on Land
The transition of vertebrates onto land required a number of changes in body design, as movement on land presents a number of challenges for animals that are adapted to movement in water. The buoyancy of water provides a certain amount of lift, and a common form of movement by fish is lateral undulations of the entire body. This back and forth movement

pushes the body against the water, creating forward movement. In most fish, the muscles of paired fins attach to girdles within the body, allowing for some control of locomotion. As certain fish began moving onto land, they retained their lateral undulation form of locomotion (anguilliform). However, instead of pushing against water, their fins or flippers became points of contact with the ground, around which they rotated their bodies. The effect of gravity and the lack of buoyancy on land meant that body weight was suspended on the limbs, leading to increased strengthening and ossification of the limbs. The effect of gravity also required changes to the axial skeleton. Lateral undulations of land animal vertebral columns cause torsional strain. A firmer, more ossified vertebral column became common in terrestrial tetrapods because it reduces strain while providing the strength needed to support the body's weight. In later tetrapods, the vertebrae began allowing for vertical motion rather than lateral flexion. Another change in the axial skeleton was the loss of a direct attachment between the pectoral girdle and the head. This reduced the jarring to the head caused by the impact of the limbs on the ground. The vertebrae of the neck also evolved to allow movement of the head independently of the body.

The appendicular skeleton of land animals is also different from aquatic animals. The shoulders attach to the pectoral girdle through muscles and connective tissue, thus reducing the jarring of the skull. Because of a lateral undulating vertebral column, in early tetrapods, the limbs were splayed out to the side and movement occurred by performing "push-ups." The vertebrae of these animals had to move side-to-side in a similar manner to fish and reptiles. This type of motion requires large muscles to move the limbs toward the midline; it was almost like walking while doing push-ups, and it is not an efficient use of energy. Later tetrapods have their limbs placed under their bodies, so that each stride requires less force to move forward. This resulted in decreased adductor muscle size and an increased range of motion of the scapulae. This also restricts movement primarily to one plane, creating forward motion rather than moving the limbs upward as well as forward. The femur and humerus were also rotated, so that the ends of the limbs and digits were pointed forward, in the direction of motion, rather than out to the side. By placement underneath the body, limbs can swing forward like a pendulum to produce a stride that is more efficient for moving over land.

Section Summary

The three types of skeleton designs are hydrostatic skeletons, exoskeletons, and endoskeletons. A hydrostatic skeleton is formed by a fluid-filled compartment held under hydrostatic pressure; movement is created by the muscles producing pressure on the fluid. An exoskeleton is a hard external skeleton that protects the outer surface of an organism and enables movement through muscles attached on the inside. An endoskeleton is an internal skeleton composed of hard, mineralized tissue that also enables movement by attachment to muscles. The human skeleton is an endoskeleton that is composed of the axial and appendicular skeleton. The axial skeleton is composed of the bones of the skull, ossicles of the ear, hyoid bone, vertebral column, and ribcage. The skull consists of eight cranial bones and 14 facial bones. Six bones make up the ossicles of the middle ear, while the hyoid bone is located in the neck under the mandible. The vertebral column contains 26 bones, and it surrounds and protects the spinal cord. The thoracic cage consists of the sternum, ribs, thoracic vertebrae, and costal cartilages. The appendicular skeleton is made up of the limbs of the upper and lower limbs. The pectoral girdle is composed of the clavicles and the scapulae. The upper limb contains 30 bones in the arm, the forearm, and the hand. The pelvic girdle attaches the lower limbs to the axial skeleton. The lower limb includes the bones of the thigh, the leg, and the foot.

Review Questions

Exercise:

Problem: The forearm consists of the:

a. radius and ulna
b. radius and humerus
c. ulna and humerus
d. humerus and carpus

Solution:

A

Exercise:

Problem: The pectoral girdle consists of the:

a. clavicle and sternum
b. sternum and scapula
c. clavicle and scapula
d. clavicle and coccyx

Solution:

C

Exercise:

Problem:

All of the following are groups of vertebrae except _________, which is a curvature.

a. thoracic
b. cervical
c. lumbar
d. pelvic

Solution:

D

Exercise:

Problem: Which of these is a facial bone?

a. frontal
b. occipital
c. lacrimal

d. temporal

Solution:

C

Free Response

Exercise:

Problem:

What are the major differences between the male pelvis and female pelvis that permit childbirth in females?

Solution:

The female pelvis is tilted forward and is wider, lighter, and shallower than the male pelvis. It is also has a pubic angle that is broader than the male pelvis.

Exercise:

Problem:

What are the major differences between the pelvic girdle and the pectoral girdle that allow the pelvic girdle to bear the weight of the body?

Solution:

The pelvic girdle is securely attached to the body by strong ligaments, unlike the pectoral girdle, which is sparingly attached to the ribcage. The sockets of the pelvic girdle are deep, allowing the femur to be more stable than the pectoral girdle, which has shallow sockets for the scapula. Most tetrapods have 75 percent of their weight on the front legs because the head and neck are so heavy; the advantage of the shoulder joint is more degrees of freedom in movement.

Glossary

appendicular skeleton
composed of the bones of the upper limbs, which function to grasp and manipulate objects, and the lower limbs, which permit locomotion

articulation
any place where two bones are joined

auditory ossicle
(also, middle ear) transduces sounds from the air into vibrations in the fluid-filled cochlea

axial skeleton
forms the central axis of the body and includes the bones of the skull, the ossicles of the middle ear, the hyoid bone of the throat, the vertebral column, and the thoracic cage (ribcage)

carpus
eight bones that comprise the wrist

clavicle
S-shaped bone that positions the arms laterally

coxal bone
hip bone

cranial bone
one of eight bones that form the cranial cavity that encloses the brain and serves as an attachment site for the muscles of the head and neck

endoskeleton
skeleton of living cells that produce a hard, mineralized tissue located within the soft tissue of organisms

exoskeleton

a secreted cellular product external skeleton that consists of a hard encasement on the surface of an organism

facial bone
: one of the 14 bones that form the face; provides cavities for the sense organs (eyes, mouth, and nose) and attachment points for facial muscles

femur
: (also, thighbone) longest, heaviest, and strongest bone in the body

fibula
: (also, calf bone) parallels and articulates with the tibia

forearm
: extends from the elbow to the wrist and consists of two bones: the ulna and the radius

humerus
: only bone of the arm

hydrostatic skeleton
: skeleton that consists of aqueous fluid held under pressure in a closed body compartment

hyoid bone
: lies below the mandible in the front of the neck

intervertebral disc
: composed of fibrous cartilage; lies between adjacent vertebrae from the second cervical vertebra to the sacrum

lower limb
: consists of the thigh, the leg, and the foot

metacarpus
: five bones that comprise the palm

metatarsal

one of the five bones of the foot

patella
(also, kneecap) triangular bone that lies anterior to the knee joint

pectoral girdle
bones that transmit the force generated by the upper limbs to the axial skeleton

phalange
one of the bones of the fingers or toes

pelvic girdle
bones that transmit the force generated by the lower limbs to the axial skeleton

radius
bone located along the lateral (thumb) side of the forearm; articulates with the humerus at the elbow

rib
one of 12 pairs of long, curved bones that attach to the thoracic vertebrae and curve toward the front of the body to form the ribcage

scapula
flat, triangular bone located at the posterior pectoral girdle

skull
bone that supports the structures of the face and protects the brain

sternum
(also, breastbone) long, flat bone located at the front of the chest

tarsal
one of the seven bones of the ankle

thoracic cage
(also, ribcage) skeleton of the chest, which consists of the ribs, thoracic vertebrae, sternum, and costal cartilages

tibia

(also, shinbone) large bone of the leg that is located directly below the knee

ulna

bone located on the medial aspect (pinky-finger side) of the forearm

vertebral column

(also, spine) surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck

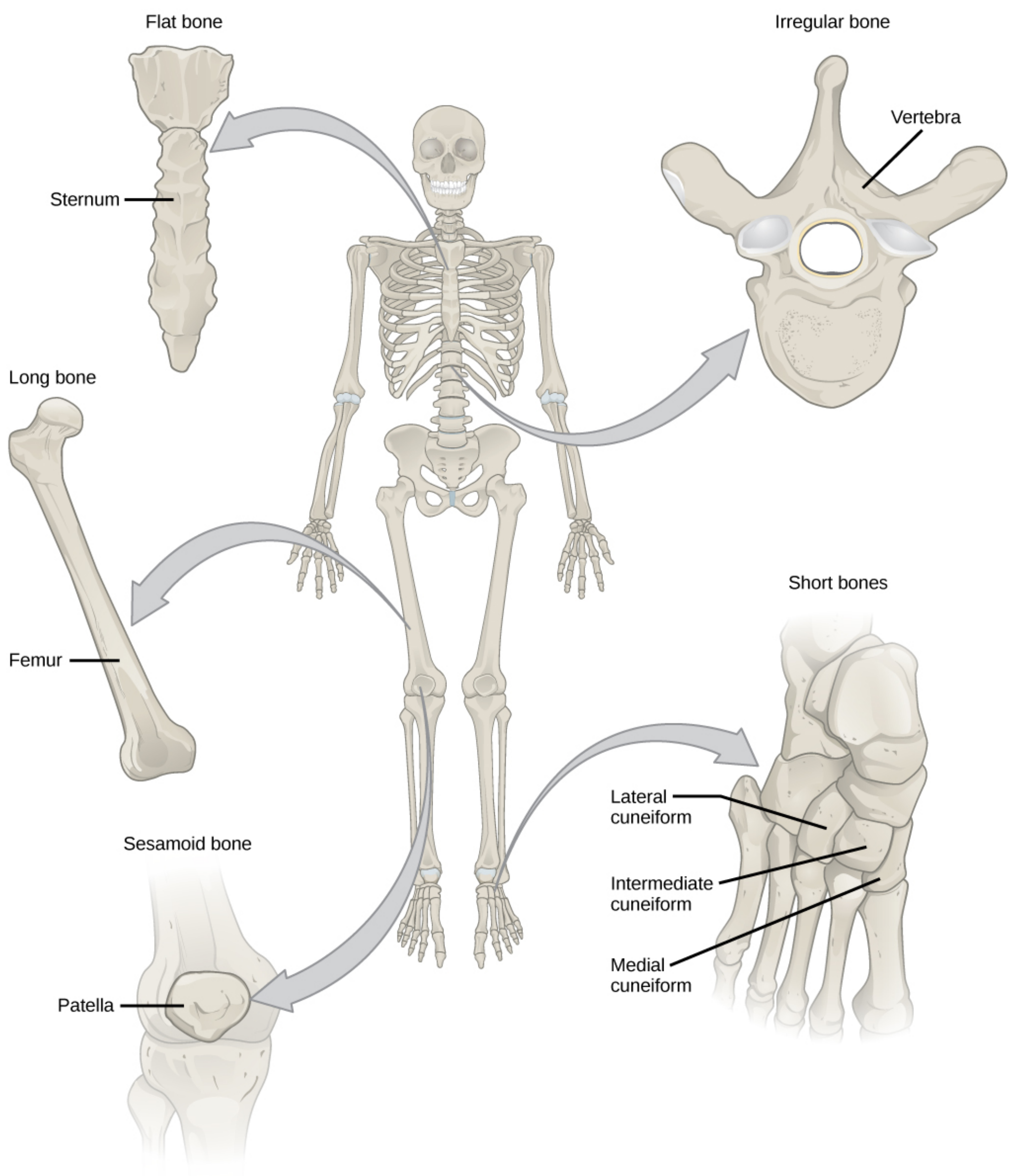
Bone
By the end of this section, you will be able to:

- Classify the different types of bones in the skeleton
- Explain the role of the different cell types in bone
- Explain how bone forms during development

Bone, or **osseous tissue**, is a connective tissue that constitutes the endoskeleton. It contains specialized cells and a matrix of mineral salts and collagen fibers.

The mineral salts primarily include hydroxyapatite, a mineral formed from calcium phosphate. **Calcification** is the process of deposition of mineral salts on the collagen fiber matrix that crystallizes and hardens the tissue. The process of calcification only occurs in the presence of collagen fibers.

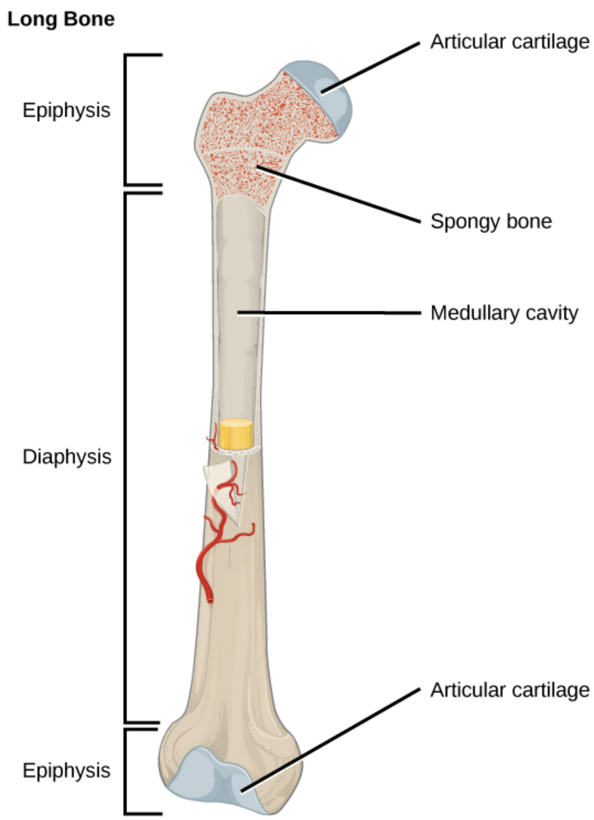
The bones of the human skeleton are classified by their shape: long bones, short bones, flat bones, sutural bones, sesamoid bones, and irregular bones ([link]).



Shown are different types of bones: flat, irregular, long, short, and sesamoid.

Long bones are longer than they are wide and have a shaft and two ends. The **diaphysis**, or central shaft, contains bone marrow in a marrow cavity. The rounded ends, the **epiphyses**, are covered with articular cartilage and are filled with red bone marrow, which produces blood cells ([link]). Most

of the limb bones are long bones—for example, the femur, tibia, ulna, and radius. Exceptions to this include the patella and the bones of the wrist and ankle.



The long bone is covered by articular cartilage at either end and contains bone marrow (shown in yellow in this illustration) in the marrow cavity.

Short bones, or cuboidal bones, are bones that are the same width and length, giving them a cube-like shape. For example, the bones of the wrist (carpals) and ankle (tarsals) are short bones ([link]).

Flat bones are thin and relatively broad bones that are found where extensive protection of organs is required or where broad surfaces of

muscle attachment are required. Examples of flat bones are the sternum (breast bone), ribs, scapulae (shoulder blades), and the roof of the skull ([link]).

Irregular bones are bones with complex shapes. These bones may have short, flat, notched, or ridged surfaces. Examples of irregular bones are the vertebrae, hip bones, and several skull bones.

Sesamoid bones are small, flat bones and are shaped similarly to a sesame seed. The patellae are sesamoid bones ([link]). Sesamoid bones develop inside tendons and may be found near joints at the knees, hands, and feet.

The patella of the knee is an example of a sesamoid bone.

Sutural bones are small, flat, irregularly shaped bones. They may be found between the flat bones of the skull. They vary in number, shape, size, and position.

Bone Tissue

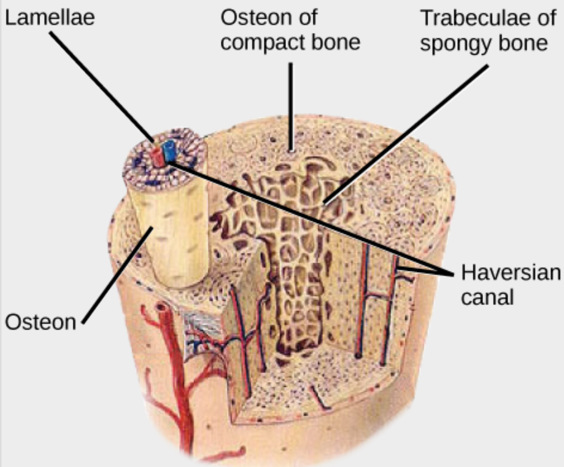
Bones are considered organs because they contain various types of tissue, such as blood, connective tissue, nerves, and bone tissue. Osteocytes, the

living cells of bone tissue, form the mineral matrix of bones. There are two types of bone tissue: compact and spongy.

Compact Bone Tissue

Compact bone (or cortical bone) forms the hard external layer of all bones and surrounds the medullary cavity, or bone marrow. It provides protection and strength to bones. Compact bone tissue consists of units called osteons or Haversian systems. **Osteons** are cylindrical structures that contain a mineral matrix and living osteocytes connected by canaliculi, which transport blood. They are aligned parallel to the long axis of the bone. Each osteon consists of **lamellae**, which are layers of compact matrix that surround a central canal called the Haversian canal. The **Haversian canal** (osteonic canal) contains the bone's blood vessels and nerve fibers ([link]). Osteons in compact bone tissue are aligned in the same direction along lines of stress and help the bone resist bending or fracturing. Therefore, compact bone tissue is prominent in areas of bone at which stresses are applied in only a few directions.

Note:
Art Connection



Compact bone tissue consists

of osteons that are aligned parallel to the long axis of the bone, and the Haversian canal that contains the bone’s blood vessels and nerve fibers. The inner layer of bones consists of spongy bone tissue. The small dark ovals in the osteon represent the living osteocytes. (credit: modification of work by NCI, NIH)

Which of the following statements about bone tissue is false?

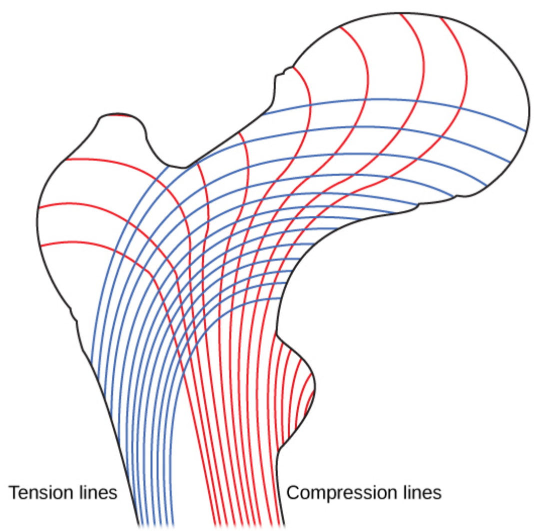
a. Compact bone tissue is made of cylindrical osteons that are aligned such that they travel the length of the bone.
b. Haversian canals contain blood vessels only.
c. Haversian canals contain blood vessels and nerve fibers.
d. Spongy tissue is found on the interior of the bone, and compact bone tissue is found on the exterior.

Spongy Bone Tissue

Whereas compact bone tissue forms the outer layer of all bones, **spongy bone** or cancellous bone forms the inner layer of all bones. Spongy bone tissue does not contain osteons that constitute compact bone tissue. Instead, it consists of **trabeculae**, which are lamellae that are arranged as rods or plates. Red bone marrow is found between the trabuculae. Blood vessels within this tissue deliver nutrients to osteocytes and remove waste. The red bone marrow of the femur and the interior of other large bones, such as the ileum, forms blood cells.

Spongy bone reduces the density of bone and allows the ends of long bones to compress as the result of stresses applied to the bone. Spongy bone is prominent in areas of bones that are not heavily stressed or where stresses arrive from many directions. The epiphyses of bones, such as the neck of the femur, are subject to stress from many directions. Imagine laying a heavy framed picture flat on the floor. You could hold up one side of the picture with a toothpick if the toothpick was perpendicular to the floor and the picture. Now drill a hole and stick the toothpick into the wall to hang up the picture. In this case, the function of the toothpick is to transmit the downward pressure of the picture to the wall. The force on the picture is straight down to the floor, but the force on the toothpick is both the picture wire pulling down and the bottom of the hole in the wall pushing up. The toothpick will break off right at the wall.

The neck of the femur is horizontal like the toothpick in the wall. The weight of the body pushes it down near the joint, but the vertical diaphysis of the femur pushes it up at the other end. The neck of the femur must be strong enough to transfer the downward force of the body weight horizontally to the vertical shaft of the femur ([link]).



Trabeculae in spongy bone are arranged such that one side of the bone bears

tension and the other
withstands compression.

Note:
Link to Learning

View micrographs of musculoskeletal tissues as you review the anatomy.

Cell Types in Bones

Bone consists of four types of cells: osteoblasts, osteoclasts, osteocytes, and osteoprogenitor cells. **Osteoblasts** are bone cells that are responsible for bone formation. Osteoblasts synthesize and secrete the organic part and inorganic part of the extracellular matrix of bone tissue, and collagen fibers. Osteoblasts become trapped in these secretions and differentiate into less active osteocytes. **Osteoclasts** are large bone cells with up to 50 nuclei. They remove bone structure by releasing lysosomal enzymes and acids that dissolve the bony matrix. These minerals, released from bones into the blood, help regulate calcium concentrations in body fluids. Bone may also be resorbed for remodeling, if the applied stresses have changed. **Osteocytes** are mature bone cells and are the main cells in bony connective tissue; these cells cannot divide. Osteocytes maintain normal bone structure by recycling the mineral salts in the bony matrix. **Osteoprogenitor cells** are squamous stem cells that divide to produce daughter cells that differentiate

into osteoblasts. Osteoprogenitor cells are important in the repair of fractures.

Development of Bone

Ossification, or osteogenesis, is the process of bone formation by osteoblasts. Ossification is distinct from the process of calcification; whereas calcification takes place during the ossification of bones, it can also occur in other tissues. Ossification begins approximately six weeks after fertilization in an embryo. Before this time, the embryonic skeleton consists entirely of fibrous membranes and hyaline cartilage. The development of bone from fibrous membranes is called intramembranous ossification; development from hyaline cartilage is called endochondral ossification. Bone growth continues until approximately age 25. Bones can grow in thickness throughout life, but after age 25, ossification functions primarily in bone remodeling and repair.

Intramembranous Ossification

Intramembranous ossification is the process of bone development from fibrous membranes. It is involved in the formation of the flat bones of the skull, the mandible, and the clavicles. Ossification begins as mesenchymal cells form a template of the future bone. They then differentiate into osteoblasts at the ossification center. Osteoblasts secrete the extracellular matrix and deposit calcium, which hardens the matrix. The non-mineralized portion of the bone or osteoid continues to form around blood vessels, forming spongy bone. Connective tissue in the matrix differentiates into red bone marrow in the fetus. The spongy bone is remodeled into a thin layer of compact bone on the surface of the spongy bone.

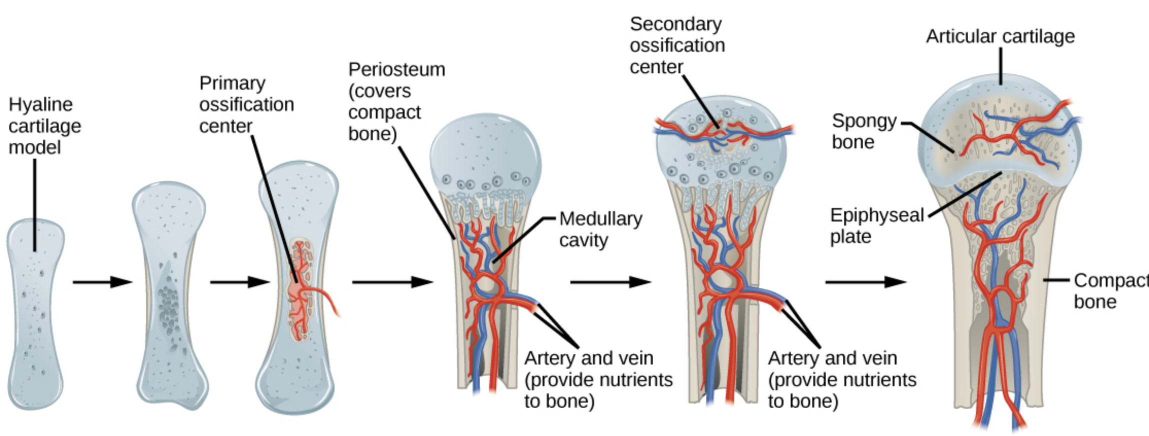
Endochondral Ossification

Endochondral ossification is the process of bone development from hyaline cartilage. All of the bones of the body, except for the flat bones of

the skull, mandible, and clavicles, are formed through endochondral ossification.

In long bones, chondrocytes form a template of the hyaline cartilage diaphysis. Responding to complex developmental signals, the matrix begins to calcify. This calcification prevents diffusion of nutrients into the matrix, resulting in chondrocytes dying and the opening up of cavities in the diaphysis cartilage. Blood vessels invade the cavities, and osteoblasts and osteoclasts modify the calcified cartilage matrix into spongy bone. Osteoclasts then break down some of the spongy bone to create a marrow, or medullary, cavity in the center of the diaphysis. Dense, irregular connective tissue forms a sheath (periosteum) around the bones. The periosteum assists in attaching the bone to surrounding tissues, tendons, and ligaments. The bone continues to grow and elongate as the cartilage cells at the epiphyses divide.

In the last stage of prenatal bone development, the centers of the epiphyses begin to calcify. Secondary ossification centers form in the epiphyses as blood vessels and osteoblasts enter these areas and convert hyaline cartilage into spongy bone. Until adolescence, hyaline cartilage persists at the **epiphyseal plate** (growth plate), which is the region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones ([link]).



Endochondral ossification is the process of bone development

from hyaline cartilage. The periosteum is the connective tissue on the outside of bone that acts as the interface between bone, blood vessels, tendons, and ligaments.

Growth of Bone

Long bones continue to lengthen, potentially until adolescence, through the addition of bone tissue at the epiphyseal plate. They also increase in width through appositional growth.

Lengthening of Long Bones

Chondrocytes on the epiphyseal side of the epiphyseal plate divide; one cell remains undifferentiated near the epiphysis, and one cell moves toward the diaphysis. The cells, which are pushed from the epiphysis, mature and are destroyed by calcification. This process replaces cartilage with bone on the diaphyseal side of the plate, resulting in a lengthening of the bone.

Long bones stop growing at around the age of 18 in females and the age of 21 in males in a process called epiphyseal plate closure. During this process, cartilage cells stop dividing and all of the cartilage is replaced by bone. The epiphyseal plate fades, leaving a structure called the epiphyseal line or epiphyseal remnant, and the epiphysis and diaphysis fuse.

Thickening of Long Bones

Appositional growth is the increase in the diameter of bones by the addition of bony tissue at the surface of bones. Osteoblasts at the bone surface secrete bone matrix, and osteoclasts on the inner surface break down bone. The osteoblasts differentiate into osteocytes. A balance between these two processes allows the bone to thicken without becoming too heavy.

Bone Remodeling and Repair

Bone renewal continues after birth into adulthood. **Bone remodeling** is the replacement of old bone tissue by new bone tissue. It involves the processes of bone deposition by osteoblasts and bone resorption by osteoclasts. Normal bone growth requires vitamins D, C, and A, plus minerals such as calcium, phosphorous, and magnesium. Hormones such as parathyroid hormone, growth hormone, and calcitonin are also required for proper bone growth and maintenance.

Bone turnover rates are quite high, with five to seven percent of bone mass being recycled every week. Differences in turnover rate exist in different areas of the skeleton and in different areas of a bone. For example, the bone in the head of the femur may be fully replaced every six months, whereas the bone along the shaft is altered much more slowly.

Bone remodeling allows bones to adapt to stresses by becoming thicker and stronger when subjected to stress. Bones that are not subject to normal stress, for example when a limb is in a cast, will begin to lose mass. A fractured or broken bone undergoes repair through four stages:

1. Blood vessels in the broken bone tear and hemorrhage, resulting in the formation of clotted blood, or a hematoma, at the site of the break. The severed blood vessels at the broken ends of the bone are sealed by the clotting process, and bone cells that are deprived of nutrients begin to die.
2. Within days of the fracture, capillaries grow into the hematoma, and phagocytic cells begin to clear away the dead cells. Though fragments of the blood clot may remain, fibroblasts and osteoblasts enter the area and begin to reform bone. Fibroblasts produce collagen fibers that connect the broken bone ends, and osteoblasts start to form spongy bone. The repair tissue between the broken bone ends is called the fibrocartilaginous callus, as it is composed of both hyaline and fibrocartilage ([link]). Some bone spicules may also appear at this point.
3. The fibrocartilaginous callus is converted into a bony callus of spongy bone. It takes about two months for the broken bone ends to be firmly

joined together after the fracture. This is similar to the endochondral formation of bone, as cartilage becomes ossified; osteoblasts, osteoclasts, and bone matrix are present.

4. The bony callus is then remodelled by osteoclasts and osteoblasts, with excess material on the exterior of the bone and within the medullary cavity being removed. Compact bone is added to create bone tissue that is similar to the original, unbroken bone. This remodeling can take many months, and the bone may remain uneven for years.

After this bone is set, a callus will knit the two ends together. (credit: Bill Rhodes)

Note:
Scientific Method Connection
Decalcification of Bones

Question: What effect does the removal of calcium and collagen have on bone structure?
Background: Conduct a literature search on the role of calcium and collagen in maintaining bone structure. Conduct a literature search on diseases in which bone structure is compromised.
Hypothesis: Develop a hypothesis that states predictions of the flexibility, strength, and mass of bones that have had the calcium and collagen components removed. Develop a hypothesis regarding the attempt to add calcium back to decalcified bones.
Test the hypothesis: Test the prediction by removing calcium from chicken bones by placing them in a jar of vinegar for seven days. Test the hypothesis regarding adding calcium back to decalcified bone by placing the decalcified chicken bones into a jar of water with calcium supplements added. Test the prediction by denaturing the collagen from the bones by baking them at 250°C for three hours.
Analyze the data: Create a table showing the changes in bone flexibility, strength, and mass in the three different environments.
Report the results: Under which conditions was the bone most flexible? Under which conditions was the bone the strongest?
Draw a conclusion: Did the results support or refute the hypothesis? How do the results observed in this experiment correspond to diseases that destroy bone tissue?

Section Summary

Bone, or osseous tissue, is connective tissue that includes specialized cells, mineral salts, and collagen fibers. The human skeleton can be divided into long bones, short bones, flat bones, and irregular bones. Compact bone tissue is composed of osteons and forms the external layer of all bones. Spongy bone tissue is composed of trabeculae and forms the inner part of all bones. Four types of cells compose bony tissue: osteocytes, osteoclasts, osteoprogenitor cells, and osteoblasts. Ossification is the process of bone formation by osteoblasts. Intramembranous ossification is the process of bone development from fibrous membranes. Endochondral ossification is the process of bone development from hyaline cartilage. Long bones

lengthen as chondrocytes divide and secrete hyaline cartilage. Osteoblasts replace cartilage with bone. Appositional growth is the increase in the diameter of bones by the addition of bone tissue at the surface of bones. Bone remodeling involves the processes of bone deposition by osteoblasts and bone resorption by osteoclasts. Bone repair occurs in four stages and can take several months.

Art Exercise

Exercise:

Problem:

[link] Which of the following statements about bone tissue is false?

a. Compact bone tissue is made of cylindrical osteons that are aligned such that they travel the length of the bone.
b. Haversian canals contain blood vessels only.
c. Haversian canals contain blood vessels and nerve fibers.
d. Spongy tissue is found on the interior of the bone, and compact bone tissue is found on the exterior.

Solution:

[link]B

Review Questions

Exercise:

Problem: The Haversian canal:

a. is arranged as rods or plates
b. contains the bone’s blood vessels and nerve fibers
c. is responsible for the lengthwise growth of long bones
d. synthesizes and secretes matrix

Solution:

B

Exercise:

Problem: The epiphyseal plate:

a. is arranged as rods or plates
b. contains the bone’s blood vessels and nerve fibers
c. is responsible for the lengthwise growth of long bones
d. synthesizes and secretes bone matrix

Solution:

C

Exercise:

Problem: The cells responsible for bone resorption are ________.

a. osteoclasts
b. osteoblasts
c. fibroblasts
d. osteocytes

Solution:

A

Exercise:

Problem: Compact bone is composed of ________.

a. trabeculae

b. compacted collagen
c. osteons
d. calcium phosphate only

Solution:

C

Free Response

Exercise:

Problem:

What are the major differences between spongy bone and compact bone?

Solution:

Compact bone tissue forms the hard external layer of all bones and consists of osteons. Compact bone tissue is prominent in areas of bone at which stresses are applied in only a few directions. Spongy bone tissue forms the inner layer of all bones and consists of trabeculae. Spongy bone is prominent in areas of bones that are not heavily stressed or at which stresses arrive from many directions.

Exercise:

Problem:

What are the roles of osteoblasts, osteocytes, and osteoclasts?

Solution:

Osteocytes function in the exchange of nutrients and wastes with the blood. They also maintain normal bone structure by recycling the mineral salts in the bony matrix. Osteoclasts remove bone tissue by

releasing lysosomal enzymes and acids that dissolve the bony matrix. Osteoblasts are bone cells that are responsible for bone formation.

Glossary

appositional growth
: increase in the diameter of bones by the addition of bone tissue at the surface of bones

bone
: (also, osseous tissue) connective tissue that constitutes the endoskeleton

bone remodeling
: replacement of old bone tissue by new bone tissue

calcification
: process of deposition of mineral salts in the collagen fiber matrix that crystallizes and hardens the tissue

compact bone
: forms the hard external layer of all bones

diaphysis
: central shaft of bone, contains bone marrow in a marrow cavity

endochondral ossification
: process of bone development from hyaline cartilage

epiphyseal plate
: region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones

epiphysis
: rounded end of bone, covered with articular cartilage and filled with red bone marrow, which produces blood cells

flat bone
: thin and relatively broad bone found where extensive protection of organs is required or where broad surfaces of muscle attachment are required

Haversian canal
: contains the bone’s blood vessels and nerve fibers

intramembranous ossification
: process of bone development from fibrous membranes

irregular bone
: bone with complex shapes; examples include vertebrae and hip bones

lamella
: layer of compact tissue that surrounds a central canal called the Haversian canal

long bone
: bone that is longer than wide, and has a shaft and two ends

osteoblast
: bone cell responsible for bone formation

osteoclast
: large bone cells with up to 50 nuclei, responsible for bone remodeling

osteocyte
: mature bone cells and the main cell in bone tissue

osseous tissue
: connective tissue that constitutes the endoskeleton

ossification
: (also, osteogenesis) process of bone formation by osteoblasts

osteon
: cylindrical structure aligned parallel to the long axis of the bone

resorption
process by which osteoclasts release minerals stored in bones

sesamoid bone
small, flat bone shaped like a sesame seed; develops inside tendons

short bone
bone that has the same width and length, giving it a cube-like shape

spongy bone tissue
forms the inner layer of all bones

suture bone
small, flat, irregularly shaped bone that forms between the flat bones of the cranium

trabeculae
lamellae that are arranged as rods or plates

Joints and Skeletal Movement
By the end of this section, you will be able to:

- Classify the different types of joints on the basis of structure
- Explain the role of joints in skeletal movement

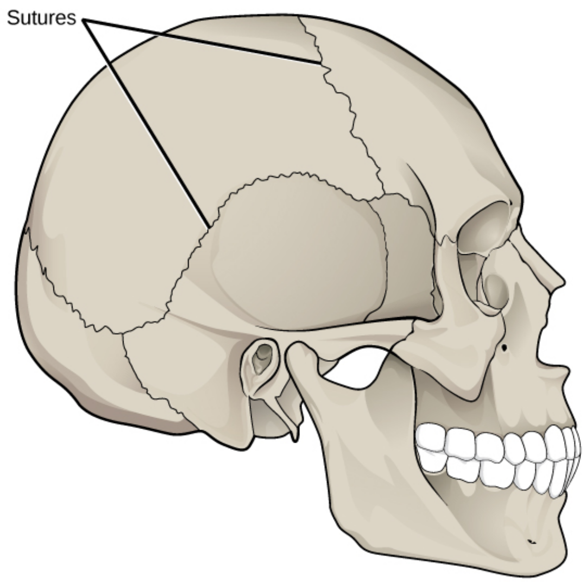
The point at which two or more bones meet is called a **joint**, or **articulation**. Joints are responsible for movement, such as the movement of limbs, and stability, such as the stability found in the bones of the skull.

Classification of Joints on the Basis of Structure

There are two ways to classify joints: on the basis of their structure or on the basis of their function. The structural classification divides joints into bony, fibrous, cartilaginous, and synovial joints depending on the material composing the joint and the presence or absence of a cavity in the joint.

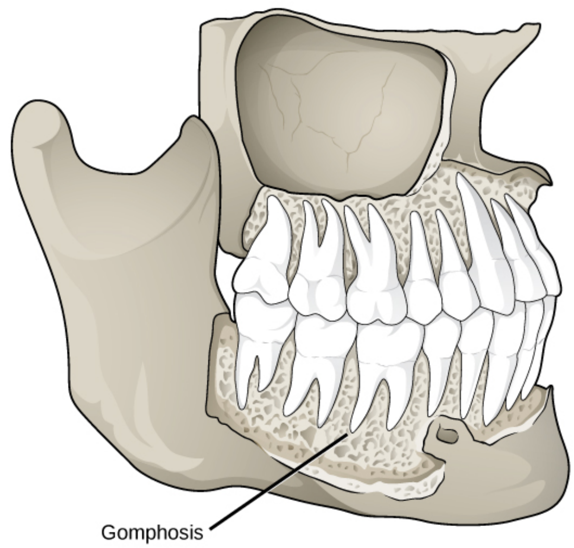
Fibrous Joints

The bones of **fibrous joints** are held together by fibrous connective tissue. There is no cavity, or space, present between the bones and so most fibrous joints do not move at all, or are only capable of minor movements. There are three types of fibrous joints: sutures, syndesmoses, and gomphoses. **Sutures** are found only in the skull and possess short fibers of connective tissue that hold the skull bones tightly in place ([link]).



Sutures are fibrous joints found only in the skull.

Syndesmoses are joints in which the bones are connected by a band of connective tissue, allowing for more movement than in a suture. An example of a syndesmosis is the joint of the tibia and fibula in the ankle. The amount of movement in these types of joints is determined by the length of the connective tissue fibers. **Gomphoses** occur between teeth and their sockets; the term refers to the way the tooth fits into the socket like a peg ([link]). The tooth is connected to the socket by a connective tissue referred to as the periodontal ligament.



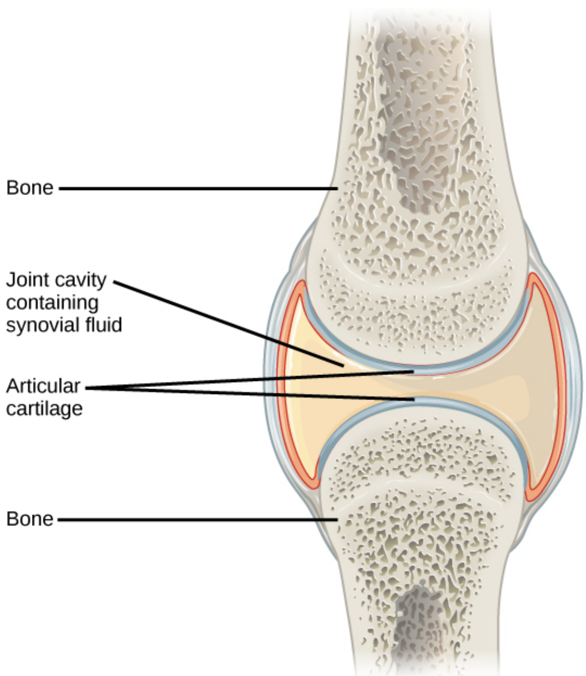
Gomphoses are fibrous joints between the teeth and their sockets. (credit: modification of work by Gray's Anatomy)

Cartilaginous Joints

Cartilaginous joints are joints in which the bones are connected by cartilage. There are two types of cartilaginous joints: synchondroses and symphyses. In a **synchondrosis**, the bones are joined by hyaline cartilage. Synchondroses are found in the epiphyseal plates of growing bones in children. In **symphyses**, hyaline cartilage covers the end of the bone but the connection between bones occurs through fibrocartilage. Symphyses are found at the joints between vertebrae. Either type of cartilaginous joint allows for very little movement.

Synovial Joints

Synovial joints are the only joints that have a space between the adjoining bones ([link]). This space is referred to as the synovial (or joint) cavity and is filled with synovial fluid. Synovial fluid lubricates the joint, reducing friction between the bones and allowing for greater movement. The ends of the bones are covered with articular cartilage, a hyaline cartilage, and the entire joint is surrounded by an articular capsule composed of connective tissue that allows movement of the joint while resisting dislocation. Articular capsules may also possess ligaments that hold the bones together. Synovial joints are capable of the greatest movement of the three structural joint types; however, the more mobile a joint, the weaker the joint. Knees, elbows, and shoulders are examples of synovial joints.



Synovial joints are the only joints that have a space or "synovial cavity" in the joint.

Classification of Joints on the Basis of Function

The functional classification divides joints into three categories: synarthroses, amphiarthroses, and diarthroses. A **synarthrosis** is a joint that is immovable. This includes sutures, gomphoses, and synchondroses. **Amphiarthroses** are joints that allow slight movement, including syndesmoses and symphyses. **Diarthroses** are joints that allow for free movement of the joint, as in synovial joints.

Movement at Synovial Joints

The wide range of movement allowed by synovial joints produces different types of movements. The movement of synovial joints can be classified as one of four different types: gliding, angular, rotational, or special movement.

Gliding Movement

Gliding movements occur as relatively flat bone surfaces move past each other. Gliding movements produce very little rotation or angular movement of the bones. The joints of the carpal and tarsal bones are examples of joints that produce gliding movements.

Angular Movement

Angular movements are produced when the angle between the bones of a joint changes. There are several different types of angular movements, including flexion, extension, hyperextension, abduction, adduction, and circumduction. **Flexion**, or bending, occurs when the angle between the bones decreases. Moving the forearm upward at the elbow or moving the wrist to move the hand toward the forearm are examples of flexion. **Extension** is the opposite of flexion in that the angle between the bones of a joint increases. Straightening a limb after flexion is an example of extension. Extension past the regular anatomical position is referred to as **hyperextension**. This includes moving the neck back to look upward, or bending the wrist so that the hand moves away from the forearm.

Abduction occurs when a bone moves away from the midline of the body. Examples of abduction are moving the arms or legs laterally to lift them straight out to the side. **Adduction** is the movement of a bone toward the midline of the body. Movement of the limbs inward after abduction is an example of adduction. **Circumduction** is the movement of a limb in a circular motion, as in moving the arm in a circular motion.

Rotational Movement

Rotational movement is the movement of a bone as it rotates around its longitudinal axis. Rotation can be toward the midline of the body, which is referred to as **medial rotation**, or away from the midline of the body, which is referred to as **lateral rotation**. Movement of the head from side to side is an example of rotation.

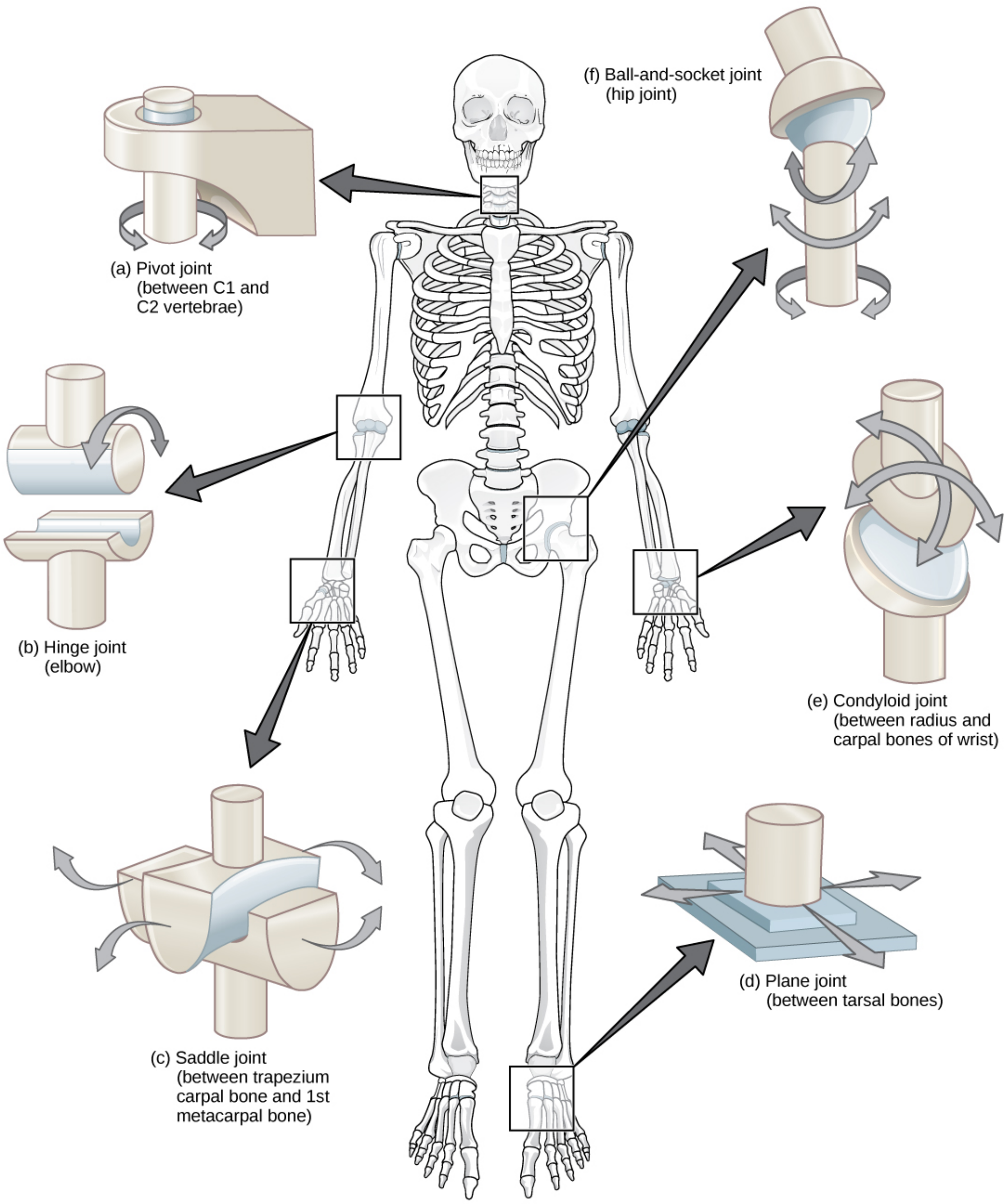
Special Movements

Some movements that cannot be classified as gliding, angular, or rotational are called special movements. **Inversion** involves the soles of the feet moving inward, toward the midline of the body. **Eversion** is the opposite of inversion, movement of the sole of the foot outward, away from the midline of the body. **Protraction** is the anterior movement of a bone in the horizontal plane. **Retraction** occurs as a joint moves back into position after protraction. Protraction and retraction can be seen in the movement of the mandible as the jaw is thrust outwards and then back inwards. **Elevation** is the movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae. **Depression** is the opposite of elevation—movement downward of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position. **Dorsiflexion** is a bending at the ankle such that the toes are lifted toward the knee. **Plantar flexion** is a bending at the ankle when the heel is lifted, such as when standing on the toes. **Supination** is the movement of the radius and ulna bones of the forearm so that the palm faces forward. **Pronation** is the opposite movement, in which the palm faces backward.

Opposition is the movement of the thumb toward the fingers of the same hand, making it possible to grasp and hold objects.

Types of Synovial Joints

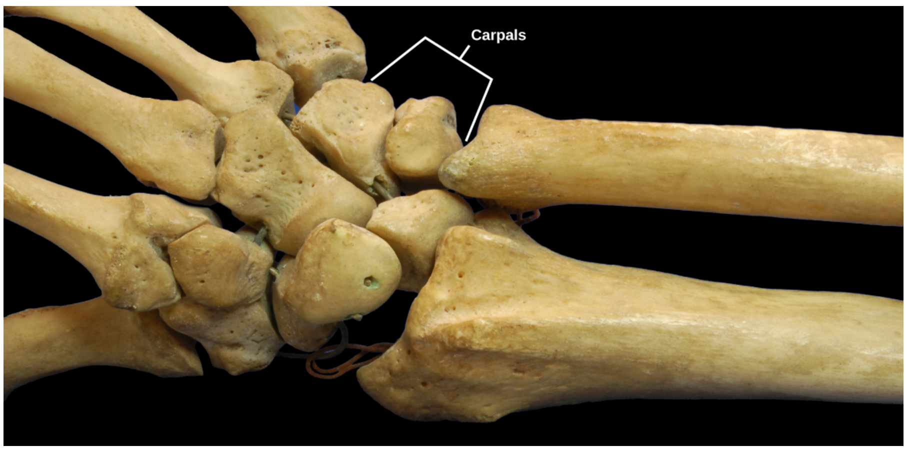
Synovial joints are further classified into six different categories on the basis of the shape and structure of the joint. The shape of the joint affects the type of movement permitted by the joint ([link]). These joints can be described as planar, hinge, pivot, condyloid, saddle, or ball-and-socket joints.



Different types of joints allow different types of movement. Planar, hinge, pivot, condyloid, saddle, and ball-and-socket are all types of synovial joints.

Planar Joints

Planar joints have bones with articulating surfaces that are flat or slightly curved faces. These joints allow for gliding movements, and so the joints are sometimes referred to as gliding joints. The range of motion is limited in these joints and does not involve rotation. Planar joints are found in the carpal bones in the hand and the tarsal bones of the foot, as well as between vertebrae ([link]).



The joints of the carpal bones in the wrist are examples of planar joints. (credit: modification of work by Brian C. Goss)

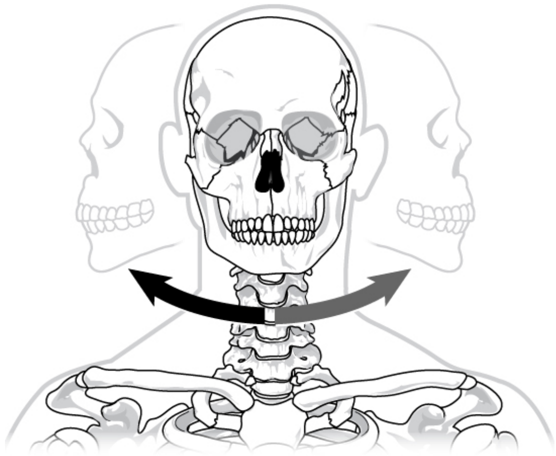
Hinge Joints

In **hinge joints**, the slightly rounded end of one bone fits into the slightly hollow end of the other bone. In this way, one bone moves while the other remains stationary, like the hinge of a door. The elbow is an example of a hinge joint. The knee is sometimes classified as a modified hinge joint ([link]).

The elbow joint, where the radius articulates with the humerus, is an example of a hinge joint. (credit: modification of work by Brian C. Goss)

Pivot Joints

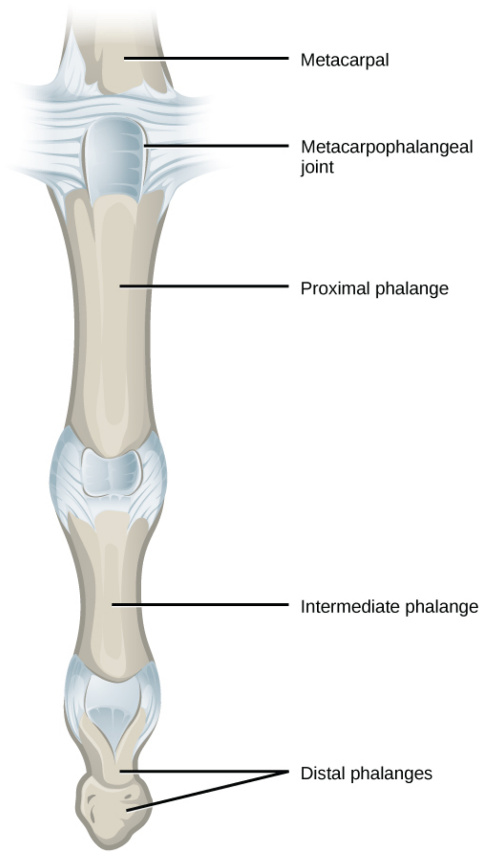
Pivot joints consist of the rounded end of one bone fitting into a ring formed by the other bone. This structure allows rotational movement, as the rounded bone moves around its own axis. An example of a pivot joint is the joint of the first and second vertebrae of the neck that allows the head to move back and forth ([link]). The joint of the wrist that allows the palm of the hand to be turned up and down is also a pivot joint.



The joint in the neck that allows the head to move back and forth is an example of a pivot joint.

Condyloid Joints

Condyloid joints consist of an oval-shaped end of one bone fitting into a similarly oval-shaped hollow of another bone ([link]). This is also sometimes called an ellipsoidal joint. This type of joint allows angular movement along two axes, as seen in the joints of the wrist and fingers, which can move both side to side and up and down.

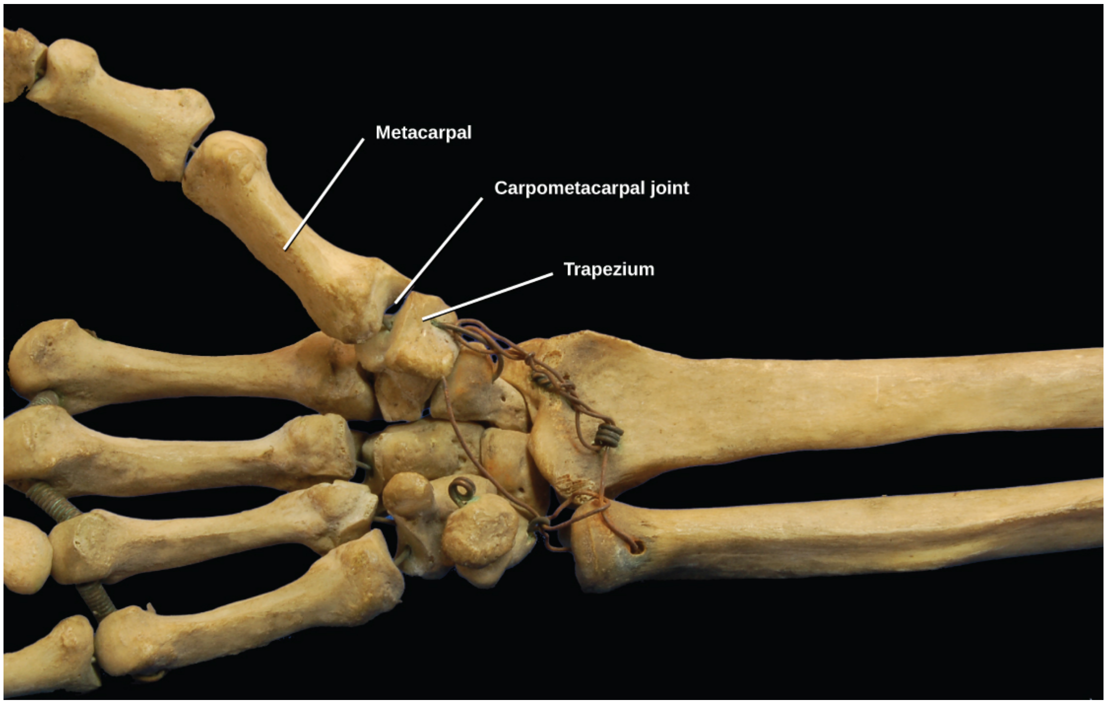


The metacarpophalangeal joints in the finger are examples of condyloid joints. (credit: modification of work by Gray's Anatomy)

Saddle Joints

Saddle joints are so named because the ends of each bone resemble a saddle, with concave and convex portions that fit together. Saddle joints allow angular movements similar to condyloid joints but with a greater range of motion. An example of a saddle joint is the thumb joint, which can

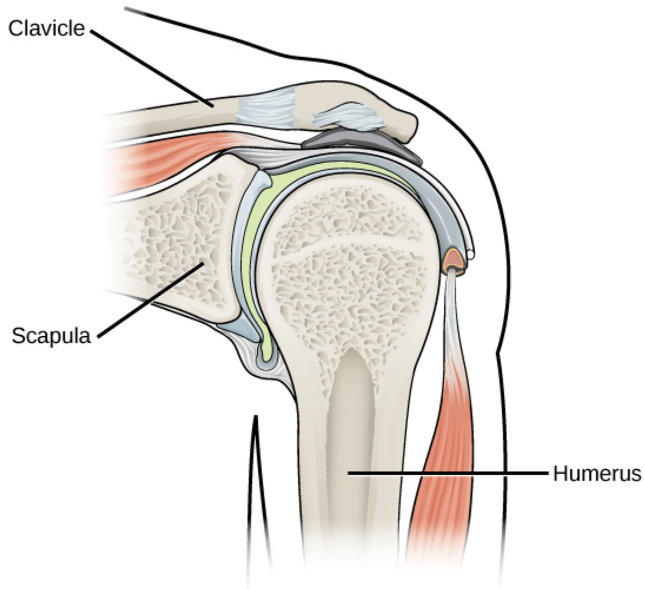
move back and forth and up and down, but more freely than the wrist or fingers ([link]).



The carpometacarpal joints in the thumb are examples of saddle joints. (credit: modification of work by Brian C. Goss)

Ball-and-Socket Joints

Ball-and-socket joints possess a rounded, ball-like end of one bone fitting into a cuplike socket of another bone. This organization allows the greatest range of motion, as all movement types are possible in all directions. Examples of ball-and-socket joints are the shoulder and hip joints ([link]).



The shoulder joint is an example of a ball-and-socket joint.

Note:
Link to Learning

Watch this animation showing the six types of synovial joints. https://www.openstaxcollege.org/l/synovial_joints

Note:
Career Connection

Rheumatologist

Rheumatologists are medical doctors who specialize in the diagnosis and treatment of disorders of the joints, muscles, and bones. They diagnose and treat diseases such as arthritis, musculoskeletal disorders, osteoporosis, and autoimmune diseases such as ankylosing spondylitis and rheumatoid arthritis.

Rheumatoid arthritis (RA) is an inflammatory disorder that primarily affects the synovial joints of the hands, feet, and cervical spine. Affected joints become swollen, stiff, and painful. Although it is known that RA is an autoimmune disease in which the body’s immune system mistakenly attacks healthy tissue, the cause of RA remains unknown. Immune cells from the blood enter joints and the synovium causing cartilage breakdown, swelling, and inflammation of the joint lining. Breakdown of cartilage causes bones to rub against each other causing pain. RA is more common in women than men and the age of onset is usually 40–50 years of age.

Rheumatologists can diagnose RA on the basis of symptoms such as joint inflammation and pain, X-ray and MRI imaging, and blood tests.

Arthrography is a type of medical imaging of joints that uses a contrast agent, such as a dye, that is opaque to X-rays. This allows the soft tissue structures of joints—such as cartilage, tendons, and ligaments—to be visualized. An arthrogram differs from a regular X-ray by showing the surface of soft tissues lining the joint in addition to joint bones. An arthrogram allows early degenerative changes in joint cartilage to be detected before bones become affected.

There is currently no cure for RA; however, rheumatologists have a number of treatment options available. Early stages can be treated with rest of the affected joints by using a cane or by using joint splints that minimize inflammation. When inflammation has decreased, exercise can be used to strengthen the muscles that surround the joint and to maintain joint flexibility. If joint damage is more extensive, medications can be used to relieve pain and decrease inflammation. Anti-inflammatory drugs such as aspirin, topical pain relievers, and corticosteroid injections may be used. Surgery may be required in cases in which joint damage is severe.

Section Summary

The structural classification of joints divides them into bony, fibrous, cartilaginous, and synovial joints. The bones of fibrous joints are held together by fibrous connective tissue; the three types of fibrous joints are sutures, syndesomes, and gomphoses. Cartilaginous joints are joints in which the bones are connected by cartilage; the two types of cartilaginous joints are synchondroses and symphyses. Synovial joints are joints that have a space between the adjoining bones. The functional classification divides joints into three categories: synarthroses, amphiarthroses, and diarthroses. The movement of synovial joints can be classified as one of four different types: gliding, angular, rotational, or special movement. Gliding movements occur as relatively flat bone surfaces move past each other. Angular movements are produced when the angle between the bones of a joint changes. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis. Special movements include inversion, eversion, protraction, retraction, elevation, depression, dorsiflexion, plantar flexion, supination, pronation, and opposition. Synovial joints are also classified into six different categories on the basis of the shape and structure of the joint: planar, hinge, pivot, condyloid, saddle, and ball-and-socket.

Review Questions

Exercise:

Problem: Synchondroses and symphyses are:

a. synovial joints
b. cartilaginous joints
c. fibrous joints
d. condyloid joints

Solution:

B

Exercise:

Problem:

The movement of bone away from the midline of the body is called ________.

a. circumduction
b. extension
c. adduction
d. abduction

Solution:

D

Exercise:

Problem:

Which of the following is not a characteristic of the synovial fluid?

a. lubrication
b. shock absorption
c. regulation of water balance in the joint
d. protection of articular cartilage

Solution:

C

Exercise:

Problem: The elbow is an example of which type of joint?

a. hinge
b. pivot
c. saddle
d. gliding

Solution:

A

Free Response

Exercise:

Problem:

What movements occur at the hip joint and knees as you bend down to touch your toes?

Solution:

The hip joint is flexed and the knees are extended.

Exercise:

Problem:

What movement(s) occur(s) at the scapulae when you shrug your shoulders?

Solution:

Elevation is the movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae. Depression is the downward movement of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position.

Glossary

abduction
: when a bone moves away from the midline of the body

adduction
movement of the limbs inward after abduction

amphiarthrosis
joint that allows slight movement; includes syndesmoses and symphyses

angular movement
produced when the angle between the bones of a joint changes

ball-and-socket joint
joint with a rounded, ball-like end of one bone fitting into a cuplike socket of another bone

cartilaginous joint
joint in which the bones are connected by cartilage

circumduction
movement of a limb in a circular motion.

condyloid joint
oval-shaped end of one bone fitting into a similarly oval-shaped hollow of another bone

depression
movement downward of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position; opposite of elevation

diarthrosis
joint that allows for free movement of the joint; found in synovial joints

dorsiflexion
bending at the ankle such that the toes are lifted toward the knee

elevation

movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae

eversion
: movement of the sole of the foot outward, away from the midline of the body; opposite of inversion

extension
: movement in which the angle between the bones of a joint increases; opposite of flexion

fibrous joint
: joint held together by fibrous connective tissue

flexion
: movement in which the angle between the bones decreases; opposite of extension

gliding movement
: when relatively flat bone surfaces move past each other

gomphosis
: the joint in which the tooth fits into the socket like a peg

hinge joint
: slightly rounded end of one bone fits into the slightly hollow end of the other bone

hyperextension
: extension past the regular anatomical position

inversion
: soles of the feet moving inward, toward the midline of the body

joint
: point at which two or more bones meet

lateral rotation
: rotation away from the midline of the body

medial rotation
: rotation toward the midline of the body

opposition
: movement of the thumb toward the fingers of the same hand, making it possible to grasp and hold objects

plantar flexion
: bending at the ankle such that the heel is lifted, such as when standing on the toes

planar joint
: joint with bones whose articulating surfaces are flat

pivot joint
: joint with the rounded end of one bone fitting into a ring formed by the other bone

pronation
: movement in which the palm faces backward

protraction
: anterior movement of a bone in the horizontal plane

retraction
: movement in which a joint moves back into position after protraction

rotational movement
: movement of a bone as it rotates around its own longitudinal axis

saddle joint
: joint with concave and convex portions that fit together; named because the ends of each bone resemble a saddle

supination
: movement of the radius and ulna bones of the forearm so that the palm faces forward

suture

short fiber of connective tissue that holds the skull bones tightly in place; found only in the skull

synarthrosis
joint that is immovable

symphysis
hyaline cartilage covers the end of the bone, but the connection between bones occurs through fibrocartilage; symphyses are found at the joints between vertebrae

synchondrosis
bones joined by hyaline cartilage; synchondroses are found in the epiphyseal plates of growing bones in children

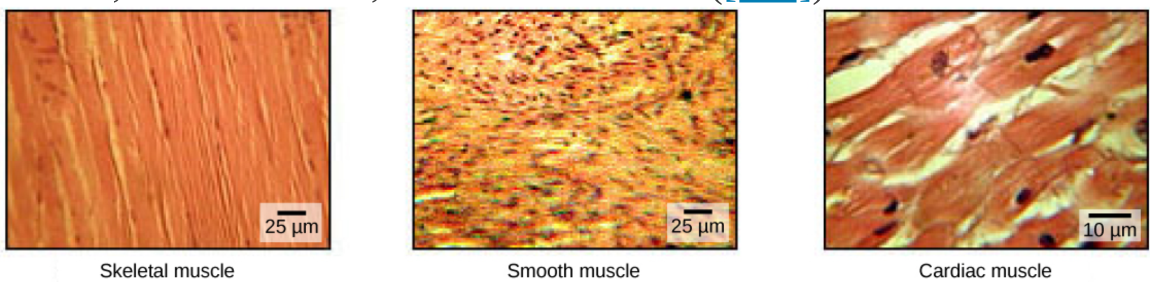
syndesmosis
joint in which the bones are connected by a band of connective tissue, allowing for more movement than in a suture

synovial joint
only joint that has a space between the adjoining bones

Muscle Contraction and Locomotion
By the end of this section, you will be able to:

- Classify the different types of muscle tissue
- Explain the role of muscles in locomotion

Muscle cells are specialized for contraction. Muscles allow for motions such as walking, and they also facilitate bodily processes such as respiration and digestion. The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle ([link]).



The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle, visualized here using light microscopy. Smooth muscle cells are short, tapered at each end, and have only one plump nucleus in each. Cardiac muscle cells are branched and striated, but short. The cytoplasm may branch, and they have one nucleus in the center of the cell. (credit: modification of work by NCI, NIH; scale-bar data from Matt Russell)

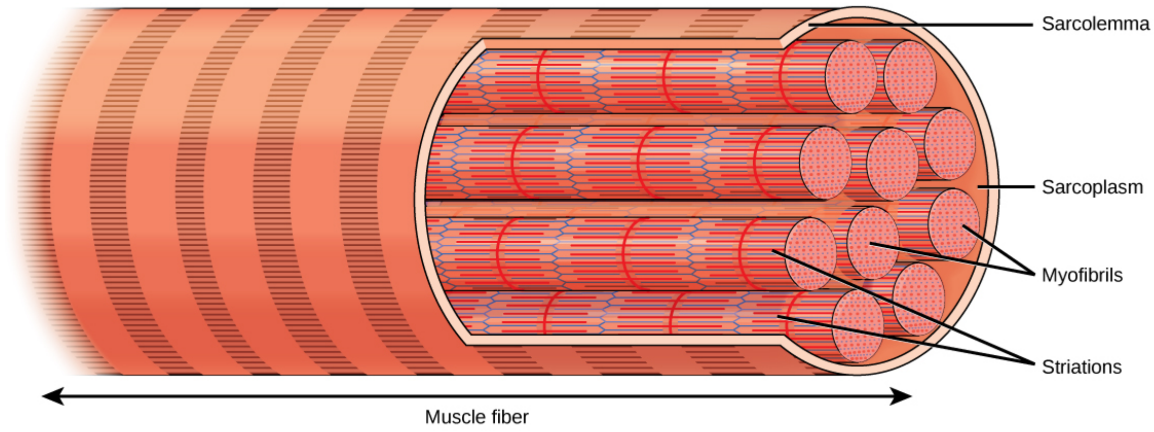
Skeletal muscle tissue forms skeletal muscles, which attach to bones or skin and control locomotion and any movement that can be consciously controlled. Because it can be controlled by thought, skeletal muscle is also called voluntary muscle. Skeletal muscles are long and cylindrical in appearance; when viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. The striations are caused by the regular arrangement of contractile proteins (actin and myosin). **Actin** is a globular contractile protein that interacts with **myosin** for muscle contraction. Skeletal muscle also has multiple nuclei present in a single cell.

Smooth muscle tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels. Smooth muscle has no striations, is not under voluntary control, has only one nucleus per cell, is tapered at both ends, and is called involuntary muscle.

Cardiac muscle tissue is only found in the heart, and cardiac contractions pump blood throughout the body and maintain blood pressure. Like skeletal muscle, cardiac muscle is striated, but unlike skeletal muscle, cardiac muscle cannot be consciously controlled and is called involuntary muscle. It has one nucleus per cell, is branched, and is distinguished by the presence of intercalated disks.

Skeletal Muscle Fiber Structure

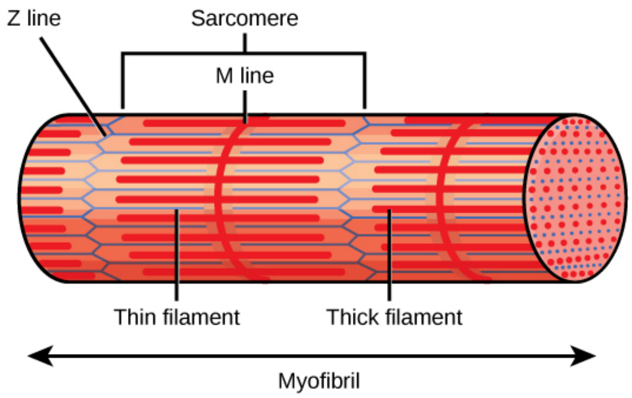
Each skeletal muscle fiber is a skeletal muscle cell. These cells are incredibly large, with diameters of up to 100 µm and lengths of up to 30 cm. The plasma membrane of a skeletal muscle fiber is called the **sarcolemma**. The sarcolemma is the site of action potential conduction, which triggers muscle contraction. Within each muscle fiber are **myofibrils**—long cylindrical structures that lie parallel to the muscle fiber. Myofibrils run the entire length of the muscle fiber, and because they are only approximately 1.2 µm in diameter, hundreds to thousands can be found inside one muscle fiber. They attach to the sarcolemma at their ends, so that as myofibrils shorten, the entire muscle cell contracts ([link]).



A skeletal muscle cell is surrounded by a plasma membrane called the sarcolemma with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils, packaged into orderly units.

The striated appearance of skeletal muscle tissue is a result of repeating bands of the proteins actin and myosin that are present along the length of myofibrils. Dark A bands and light I bands repeat along myofibrils, and the alignment of myofibrils in the cell causes the entire cell to appear striated or banded.

Each I band has a dense line running vertically through the middle called a Z disc or Z line. The Z discs mark the border of units called **sarcomeres**, which are the functional units of skeletal muscle. One sarcomere is the space between two consecutive Z discs and contains one entire A band and two halves of an I band, one on either side of the A band. A myofibril is composed of many sarcomeres running along its length, and as the sarcomeres individually contract, the myofibrils and muscle cells shorten ([link]).



A sarcomere is the region from one Z line to the next Z line. Many sarcomeres are present in a myofibril, resulting in the striation pattern characteristic of skeletal muscle.

Myofibrils are composed of smaller structures called **myofilaments**. There are two main types of filaments: thick filaments and thin filaments; each has different compositions and locations. **Thick filaments** occur only in the A band of a myofibril. **Thin filaments** attach to a protein in the Z disc called alpha-actinin and occur across the entire length of the I band and partway into the A band. The region at which thick and thin filaments overlap has a dense appearance, as there is little space between the filaments. Thin filaments do not extend all the way into the A bands, leaving a central region of the A band that only contains thick filaments. This central region of the A band looks slightly lighter than the rest of the A band and is called the H zone. The middle of the H zone has a vertical line called the M line, at which accessory proteins hold together thick filaments. Both the Z disc and the M line hold myofilaments in place to maintain the structural arrangement and layering of the myofibril. Myofibrils are connected to each other by intermediate, or desmin, filaments that attach to the Z disc.

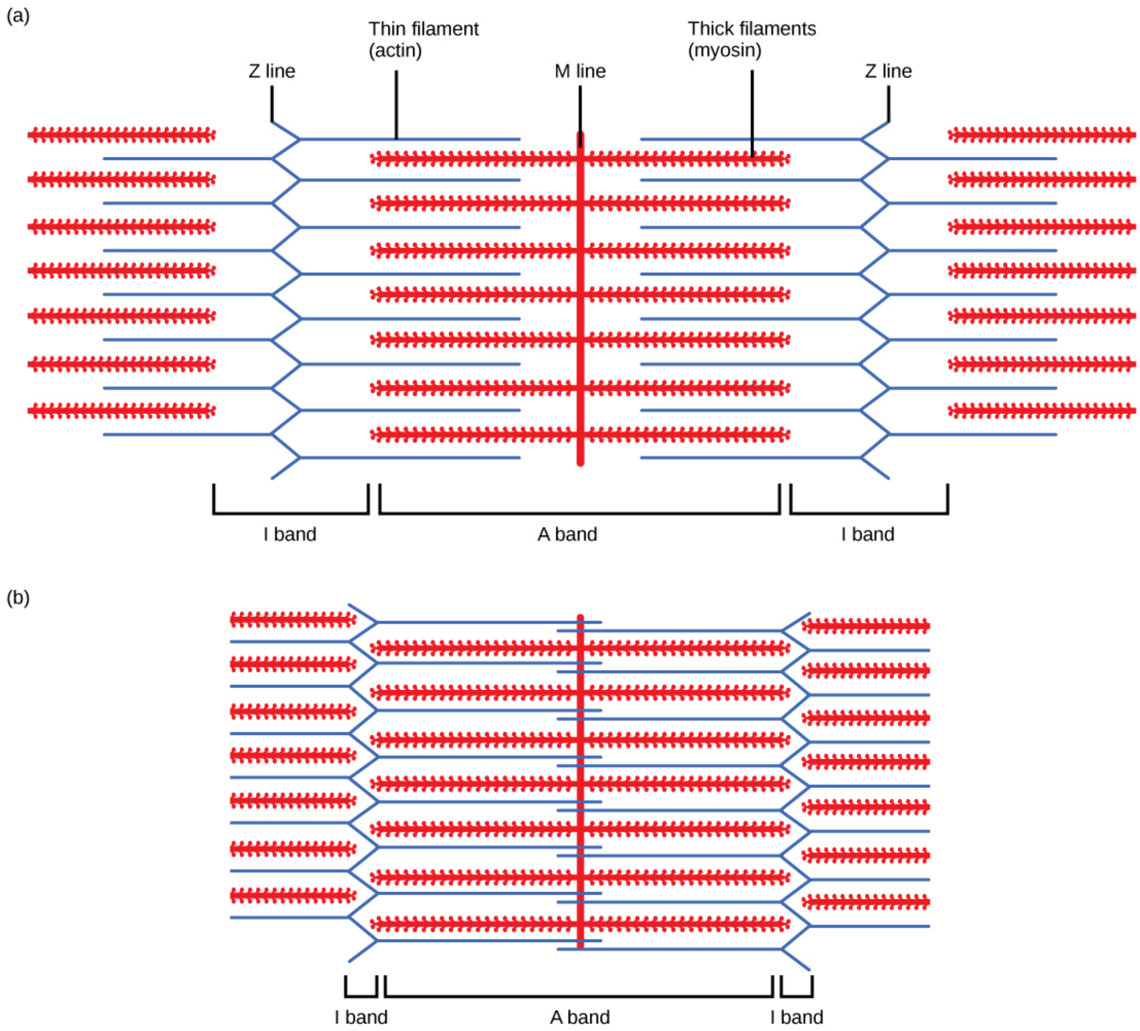
Thick and thin filaments are themselves composed of proteins. Thick filaments are composed of the protein myosin. The tail of a myosin molecule connects with other myosin molecules to form the central region of a thick filament near the M line, whereas the heads align on either side of the thick filament where the thin filaments overlap. The primary component of thin filaments is the actin protein. Two other components of the thin filament are tropomyosin and troponin. Actin has binding sites for myosin attachment. Strands of tropomyosin block the binding sites and prevent actin–myosin interactions when the muscles are at rest. Troponin consists of three globular subunits. One subunit binds to tropomyosin, one subunit binds to actin, and one subunit binds Ca^{2+} ions.

Note:
Link to Learning

View this animation showing the organization of muscle fibers.
https://www.openstaxcollege.org/l/skeletal_muscle

Sliding Filament Model of Contraction

For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments—the components of sarcomeres—do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding filament theory of muscle contraction was developed to fit the differences observed in the named bands on the sarcomere at different degrees of muscle contraction and relaxation. The mechanism of contraction is the binding of myosin to actin, forming cross-bridges that generate filament movement ([link]).



When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

When a sarcomere shortens, some regions shorten whereas others stay the same length. A sarcomere is defined as the distance between two consecutive Z discs or Z lines; when a muscle contracts, the distance between the Z discs is reduced. The H zone—the central region of the A zone—contains only thick filaments and is shortened during contraction. The I band contains only thin filaments and also shortens. The A band does not shorten—it remains the same length—but A bands of different sarcomeres move closer together during contraction, eventually

disappearing. Thin filaments are pulled by the thick filaments toward the center of the sarcomere until the Z discs approach the thick filaments. The zone of overlap, in which thin filaments and thick filaments occupy the same area, increases as the thin filaments move inward.

ATP and Muscle Contraction

The motion of muscle shortening occurs as myosin heads bind to actin and pull the actin inwards. This action requires energy, which is provided by ATP. Myosin binds to actin at a binding site on the globular actin protein. Myosin has another binding site for ATP at which enzymatic activity hydrolyzes ATP to ADP, releasing an inorganic phosphate molecule and energy.

ATP binding causes myosin to release actin, allowing actin and myosin to detach from each other. After this happens, the newly bound ATP is converted to ADP and inorganic phosphate, P_i. The enzyme at the binding site on myosin is called ATPase. The energy released during ATP hydrolysis changes the angle of the myosin head into a “cocked” position. The myosin head is then in a position for further movement, possessing potential energy, but ADP and P_i are still attached. If actin binding sites are covered and unavailable, the myosin will remain in the high energy configuration with ATP hydrolyzed, but still attached.

If the actin binding sites are uncovered, a cross-bridge will form; that is, the myosin head spans the distance between the actin and myosin molecules. P_i is then released, allowing myosin to expend the stored energy as a conformational change. The myosin head moves toward the M line, pulling the actin along with it. As the actin is pulled, the filaments move approximately 10 nm toward the M line. This movement is called the power stroke, as it is the step at which force is produced. As the actin is pulled toward the M line, the sarcomere shortens and the muscle contracts.

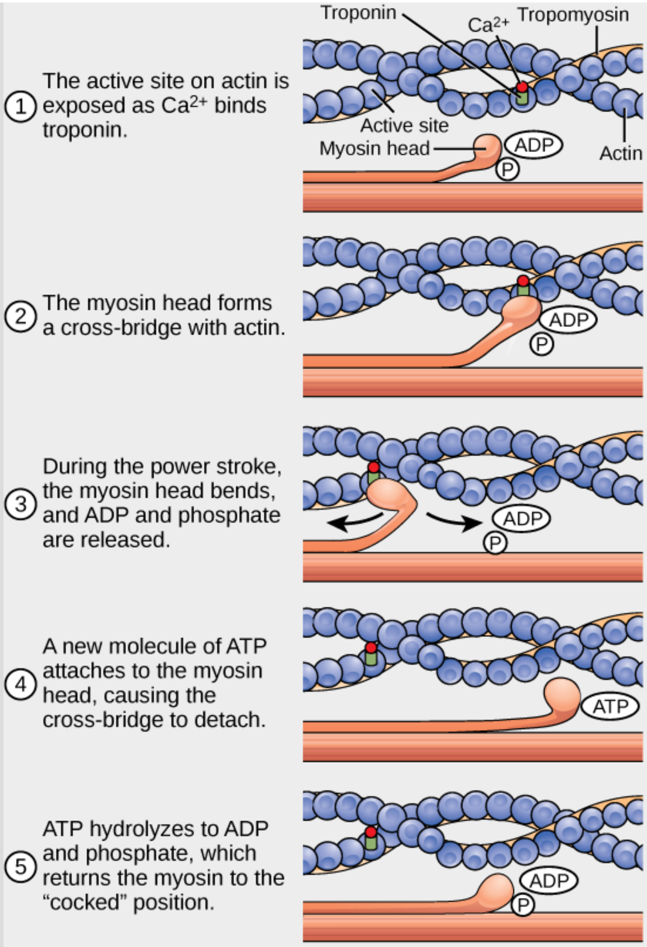
When the myosin head is “cocked,” it contains energy and is in a high-energy configuration. This energy is expended as the myosin head moves through the power stroke; at the end of the power stroke, the myosin head is in a low-energy position. After the power stroke, ADP is released; however,

the cross-bridge formed is still in place, and actin and myosin are bound together. ATP can then attach to myosin, which allows the cross-bridge cycle to start again and further muscle contraction can occur ([link]).

Note:
Link to Learning

Watch this video explaining how a muscle contraction is signaled. https://www.openstaxcollege.org/l/contract_muscle

Note:
Art Connection



The cross-bridge muscle contraction cycle, which is triggered by Ca^{2+} binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin.

Which of the following statements about muscle contraction is true?

a. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
b. The power stroke occurs when ADP and phosphate dissociate from the myosin head.

c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
d. The power stroke occurs when Ca^{2+} binds the calcium head.

Note:
Link to Learning

View this animation of the cross-bridge muscle contraction.

Regulatory Proteins

When a muscle is in a resting state, actin and myosin are separated. To keep actin from binding to the active site on myosin, regulatory proteins block the molecular binding sites. **Tropomyosin** blocks myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction in a muscle without nervous input. **Troponin** binds to tropomyosin and helps to position it on the actin molecule; it also binds calcium ions.

To enable a muscle contraction, tropomyosin must change conformation, uncovering the myosin-binding site on an actin molecule and allowing cross-bridge formation. This can only happen in the presence of calcium, which is kept at extremely low concentrations in the sarcoplasm. If present, calcium ions bind to troponin, causing conformational changes in troponin that allow tropomyosin to move away from the myosin binding sites on actin. Once the tropomyosin is removed, a cross-bridge can form between actin and myosin, triggering contraction. Cross-bridge cycling continues

until Ca^{2+} ions and ATP are no longer available and tropomyosin again covers the binding sites on actin.

Excitation–Contraction Coupling

Excitation–contraction coupling is the link (transduction) between the action potential generated in the sarcolemma and the start of a muscle contraction. The trigger for calcium release from the sarcoplasmic reticulum into the sarcoplasm is a neural signal. Each skeletal muscle fiber is controlled by a motor neuron, which conducts signals from the brain or spinal cord to the muscle. The area of the sarcolemma on the muscle fiber that interacts with the neuron is called the **motor end plate**. The end of the neuron's axon is called the synaptic terminal, and it does not actually contact the motor end plate. A small space called the synaptic cleft separates the synaptic terminal from the motor end plate. Electrical signals travel along the neuron's axon, which branches through the muscle and connects to individual muscle fibers at a neuromuscular junction.

The ability of cells to communicate electrically requires that the cells expend energy to create an electrical gradient across their cell membranes. This charge gradient is carried by ions, which are differentially distributed across the membrane. Each ion exerts an electrical influence and a concentration influence. Just as milk will eventually mix with coffee without the need to stir, ions also distribute themselves evenly, if they are permitted to do so. In this case, they are not permitted to return to an evenly mixed state.

The sodium–potassium ATPase uses cellular energy to move K^+ ions inside the cell and Na^+ ions outside. This alone accumulates a small electrical charge, but a big concentration gradient. There is lots of K^+ in the cell and lots of Na^+ outside the cell. Potassium is able to leave the cell through K^+ channels that are open 90% of the time, and it does. However, Na^+ channels are rarely open, so Na^+ remains outside the cell. When K^+ leaves the cell, obeying its concentration gradient, that effectively leaves a negative charge behind. So at rest, there is a large concentration gradient for Na^+ to enter the cell, and there is an accumulation of negative charges left behind in the cell. This is the resting membrane potential. Potential in this context means a

separation of electrical charge that is capable of doing work. It is measured in volts, just like a battery. However, the transmembrane potential is considerably smaller (0.07 V); therefore, the small value is expressed as millivolts (mV) or 70 mV. Because the inside of a cell is negative compared with the outside, a minus sign signifies the excess of negative charges inside the cell, −70 mV.

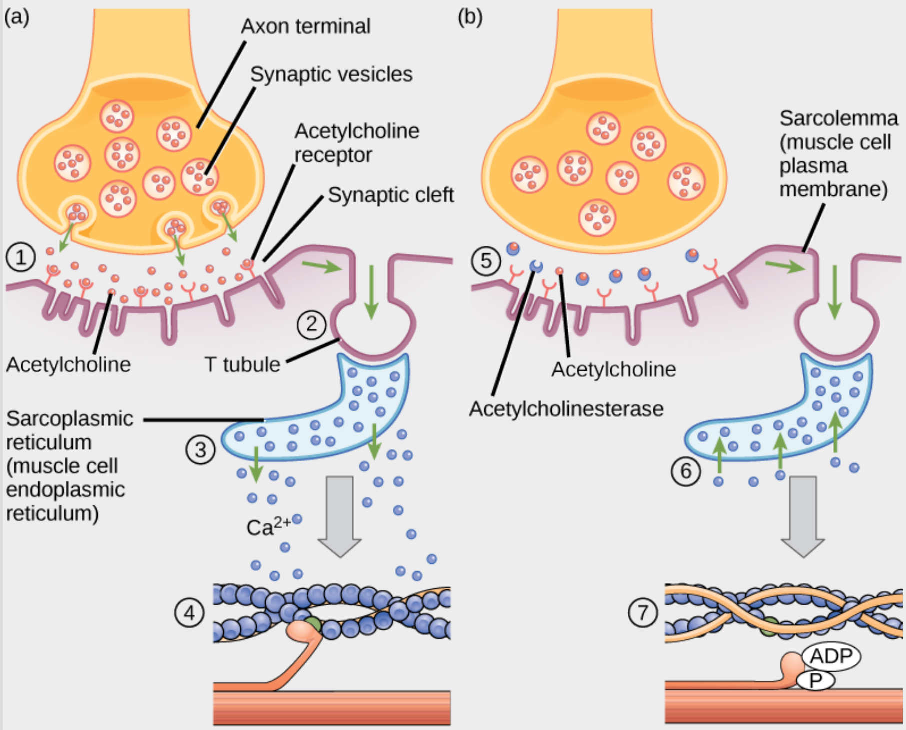
If an event changes the permeability of the membrane to Na^+ ions, they will enter the cell. That will change the voltage. This is an electrical event, called an action potential, that can be used as a cellular signal. Communication occurs between nerves and muscles through neurotransmitters. Neuron action potentials cause the release of neurotransmitters from the synaptic terminal into the synaptic cleft, where they can then diffuse across the synaptic cleft and bind to a receptor molecule on the motor end plate. The motor end plate possesses junctional folds—folds in the sarcolemma that create a large surface area for the neurotransmitter to bind to receptors. The receptors are actually sodium channels that open to allow the passage of Na^+ into the cell when they receive neurotransmitter signal.

Acetylcholine (ACh) is a neurotransmitter released by motor neurons that binds to receptors in the motor end plate. Neurotransmitter release occurs when an action potential travels down the motor neuron’s axon, resulting in altered permeability of the synaptic terminal membrane and an influx of calcium. The Ca^{2+} ions allow synaptic vesicles to move to and bind with the presynaptic membrane (on the neuron), and release neurotransmitter from the vesicles into the synaptic cleft. Once released by the synaptic terminal, ACh diffuses across the synaptic cleft to the motor end plate, where it binds with ACh receptors. As a neurotransmitter binds, these ion channels open, and Na^+ ions cross the membrane into the muscle cell. This reduces the voltage difference between the inside and outside of the cell, which is called depolarization. As ACh binds at the motor end plate, this depolarization is called an end-plate potential. The depolarization then spreads along the sarcolemma, creating an action potential as sodium channels adjacent to the initial depolarization site sense the change in voltage and open. The action potential moves across the entire cell, creating a wave of depolarization.

ACh is broken down by the enzyme **acetylcholinesterase** (AChE) into acetyl and choline. AChE resides in the synaptic cleft, breaking down ACh so that it does not remain bound to ACh receptors, which would cause unwanted extended muscle contraction ([link]).

Note:

Art Connection



1. Acetylcholine released from the axon terminal binds to receptors on the sarcolemma.
2. An action potential is generated and travels down the T tubule.
3. Ca^{2+} is released from the sarcoplasmic reticulum in response to the change in voltage.
4. Ca^{2+} binds troponin; Cross-bridges form between actin and myosin.
5. Acetylcholinesterase removes acetylcholine from the synaptic cleft.
6. Ca^{2+} is transported back into the sarcoplasmic reticulum.
7. Tropomyosin binds active sites on actin causing the cross-bridge to detach.

This diagram shows excitation-contraction coupling in a skeletal muscle contraction. The sarcoplasmic reticulum is a specialized endoplasmic reticulum found in muscle cells.

The deadly nerve gas Sarin irreversibly inhibits acetycholinesterase. What effect would Sarin have on muscle contraction?

After depolarization, the membrane returns to its resting state. This is called repolarization, during which voltage-gated sodium channels close. Potassium channels continue at 90% conductance. Because the plasma membrane sodium–potassium ATPase always transports ions, the resting state (negatively charged inside relative to the outside) is restored. The period immediately following the transmission of an impulse in a nerve or muscle, in which a neuron or muscle cell regains its ability to transmit another impulse, is called the refractory period. During the refractory period, the membrane cannot generate another action potential. . The refractory period allows the voltage-sensitive ion channels to return to their resting configurations. The sodium potassium ATPase continually moves Na^+ back out of the cell and K^+ back into the cell, and the K^+ leaks out leaving negative charge behind. Very quickly, the membrane repolarizes, so that it can again be depolarized.

Control of Muscle Tension

Neural control initiates the formation of actin–myosin cross-bridges, leading to the sarcomere shortening involved in muscle contraction. These contractions extend from the muscle fiber through connective tissue to pull on bones, causing skeletal movement. The pull exerted by a muscle is called tension, and the amount of force created by this tension can vary. This enables the same muscles to move very light objects and very heavy objects. In individual muscle fibers, the amount of tension produced depends on the cross-sectional area of the muscle fiber and the frequency of neural stimulation.

The number of cross-bridges formed between actin and myosin determine the amount of tension that a muscle fiber can produce. Cross-bridges can only form where thick and thin filaments overlap, allowing myosin to bind to actin. If more cross-bridges are formed, more myosin will pull on actin, and more tension will be produced.

The ideal length of a sarcomere during production of maximal tension occurs when thick and thin filaments overlap to the greatest degree. If a sarcomere at rest is stretched past an ideal resting length, thick and thin filaments do not overlap to the greatest degree, and fewer cross-bridges can form. This results in fewer myosin heads pulling on actin, and less tension is produced. As a sarcomere is shortened, the zone of overlap is reduced as the thin filaments reach the H zone, which is composed of myosin tails. Because it is myosin heads that form cross-bridges, actin will not bind to myosin in this zone, reducing the tension produced by this myofiber. If the sarcomere is shortened even more, thin filaments begin to overlap with each other—reducing cross-bridge formation even further, and producing even less tension. Conversely, if the sarcomere is stretched to the point at which thick and thin filaments do not overlap at all, no cross-bridges are formed and no tension is produced. This amount of stretching does not usually occur because accessory proteins, internal sensory nerves, and connective tissue oppose extreme stretching.

The primary variable determining force production is the number of myofibers within the muscle that receive an action potential from the neuron that controls that fiber. When using the biceps to pick up a pencil, the motor cortex of the brain only signals a few neurons of the biceps, and only a few myofibers respond. In vertebrates, each myofiber responds fully if stimulated. When picking up a piano, the motor cortex signals all of the neurons in the biceps and every myofiber participates. This is close to the maximum force the muscle can produce. As mentioned above, increasing the frequency of action potentials (the number of signals per second) can increase the force a bit more, because the tropomyosin is flooded with calcium.

Section Summary

The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle. Skeleton muscle tissue is composed of sarcomeres, the functional units of muscle tissue. Muscle contraction occurs when sarcomeres shorten, as thick and thin filaments slide past each other, which is called the sliding filament model of muscle contraction. ATP provides the energy for cross-bridge formation and filament sliding.

Regulatory proteins, such as troponin and tropomyosin, control cross-bridge formation. Excitation–contraction coupling transduces the electrical signal of the neuron, via acetylcholine, to an electrical signal on the muscle membrane, which initiates force production. The number of muscle fibers contracting determines how much force the whole muscle produces.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about muscle contraction is true?

a. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
b. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
d. The power stroke occurs when Ca^{2+} binds the calcium head.

Solution:

[link] B

Exercise:

Problem:

[link] The deadly nerve gas Sarin irreversibly inhibits acetycholinesterase. What effect would Sarin have on muscle contraction?

Solution:

[link] In the presence of Sarin, acetycholine is not removed from the synapse, resulting in continuous stimulation of the muscle plasma membrane. At first, muscle activity is intense and uncontrolled, but the ion gradients dissipate, so electrical signals in the T-tubules are no longer possible. The result is paralysis, leading to death by asphyxiation.

Review Questions

Exercise:

Problem:

In relaxed muscle, the myosin-binding site on actin is blocked by _________.

a. titin
b. troponin
c. myoglobin
d. tropomyosin

Solution:

D

Exercise:

Problem: The cell membrane of a muscle fiber is called a _________.

a. myofibril
b. sarcolemma
c. sarcoplasm
d. myofilament

Solution:

B

Exercise:

Problem:

The muscle relaxes if no new nerve signal arrives. However the neurotransmitter from the previous stimulation is still present in the synapse. The activity of _________ helps to remove this neurotransmitter.

a. myosin
b. action potential
c. tropomyosin
d. acetylcholinesterase

Solution:

D

Exercise:

Problem:

The ability of a muscle to generate tension immediately after stimulation is dependent on:

a. myosin interaction with the M line
b. overlap of myosin and actin
c. actin attachments to the Z line
d. none of the above

Solution:

D

Free Response

Exercise:

Problem:

How would muscle contractions be affected if ATP was completely depleted in a muscle fiber?

Solution:

Because ATP is required for myosin to release from actin, muscles would remain rigidly contracted until more ATP was available for the myosin cross-bridge release. This is why dead vertebrates undergo rigor mortis.

Exercise:

Problem:

What factors contribute to the amount of tension produced in an individual muscle fiber?

Solution:

The cross-sectional area, the length of the muscle fiber at rest, and the frequency of neural stimulation.

Exercise:

Problem:

What effect will low blood calcium have on neurons? What effect will low blood calcium have on skeletal muscles?

Solution:

Neurons will not be able to release neurotransmitter without calcium. Skeletal muscles have calcium stored and don’t need any from the outside.

Glossary

actin
: globular contractile protein that interacts with myosin for muscle contraction

acetylcholinesterase
: (AChE) enzyme that breaks down ACh into acetyl and choline

cardiac muscle
: tissue muscle tissue found only in the heart; cardiac contractions pump blood throughout the body and maintain blood pressure

motor end plate
: sarcolemma of the muscle fiber that interacts with the neuron

myofibril
: long cylindrical structures that lie parallel to the muscle fiber

myofilament
: small structures that make up myofibrils

myosin
: contractile protein that interacts with actin for muscle contraction

sarcolemma
: plasma membrane of a skeletal muscle fiber

sarcomere
: functional unit of skeletal muscle

skeletal muscle tissue
: forms skeletal muscles, which attach to bones and control locomotion and any movement that can be consciously controlled

smooth muscle
: tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels

thick filament

a group of myosin molecules

thin filament
two polymers of actin wound together along with tropomyosin and troponin

tropomyosin
acts to block myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction until a muscle receives a neuron signal

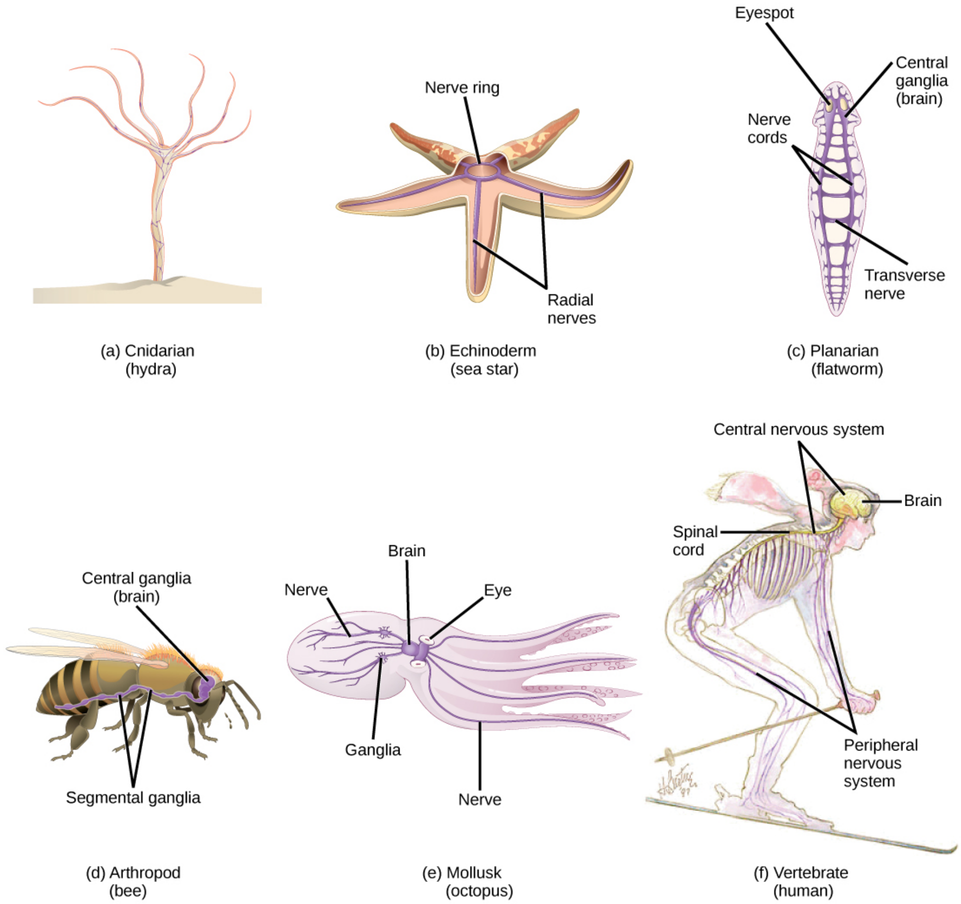
troponin
binds to tropomyosin and helps to position it on the actin molecule, and also binds calcium ions

Derived copy of Neurons and Glial Cells

By the end of this section, you will be able to:

- List and describe the functions of the structural components of a neuron
- List and describe the four main types of neurons
- Compare the functions of different types of glial cells

Nervous systems throughout the animal kingdom vary in structure and complexity, as illustrated by the variety of animals shown in [link]. Some organisms, like sea sponges, lack a true nervous system. Others, like jellyfish, lack a true brain and instead have a system of separate but connected nerve cells (neurons) called a "nerve net." Echinoderms such as sea stars have nerve cells that are bundled into fibers called nerves. Flatworms of the phylum Platyhelminthes have both a central nervous system (CNS), made up of a small "brain" and two nerve cords, and a peripheral nervous system (PNS) containing a system of nerves that extend throughout the body. The insect nervous system is more complex but also fairly decentralized. It contains a brain, ventral nerve cord, and ganglia (clusters of connected neurons). These ganglia can control movements and behaviors without input from the brain. Octopi may have the most complicated of invertebrate nervous systems—they have neurons that are organized in specialized lobes and eyes that are structurally similar to vertebrate species.



Nervous systems vary in structure and complexity. In (a) cnidarians, nerve cells form a decentralized nerve net. In (b) echinoderms, nerve cells are bundled into fibers called nerves. In animals exhibiting bilateral symmetry such as (c) planarians, neurons cluster into an anterior brain that processes information. In addition to a brain, (d) arthropods have clusters of nerve cell bodies, called peripheral ganglia, located along the ventral nerve cord. Mollusks such as squid and (e) octopi, which must hunt to survive, have complex brains containing millions of neurons. In (f) vertebrates, the brain and spinal cord comprise the central nervous system, while neurons extending into the rest of the body comprise the peripheral nervous system. (credit e: modification of work by Michael

Vecchione, Clyde F.E. Roper, and Michael J. Sweeney, NOAA; credit f: modification of work by NIH)

Compared to invertebrates, vertebrate nervous systems are more complex, centralized, and specialized. While there is great diversity among different vertebrate nervous systems, they all share a basic structure: a CNS that contains a brain and spinal cord and a PNS made up of peripheral sensory and motor nerves. One interesting difference between the nervous systems of invertebrates and vertebrates is that the nerve cords of many invertebrates are located ventrally whereas the vertebrate spinal cords are located dorsally. There is debate among evolutionary biologists as to whether these different nervous system plans evolved separately or whether the invertebrate body plan arrangement somehow “flipped” during the evolution of vertebrates.

Note:
Link to Learning

Watch this video of biologist Mark Kirschner discussing the “flipping” phenomenon of vertebrate evolution.

The nervous system is made up of **neurons**, specialized cells that can receive and transmit chemical or electrical signals, and **glia**, cells that provide support functions for the neurons by playing an information processing role that is complementary to neurons. A neuron can be compared to an electrical wire—it transmits a signal from one place to another. Glia can be compared to the workers at the electric company who

make sure wires go to the right places, maintain the wires, and take down wires that are broken. Although glia have been compared to workers, recent evidence suggests that also usurp some of the signaling functions of neurons.

There is great diversity in the types of neurons and glia that are present in different parts of the nervous system. There are four major types of neurons, and they share several important cellular components.

Neurons

The nervous system of the common laboratory fly, *Drosophila melanogaster*, contains around 100,000 neurons, the same number as a lobster. This number compares to 75 million in the mouse and 300 million in the octopus. A human brain contains around 86 billion neurons. Despite these very different numbers, the nervous systems of these animals control many of the same behaviors—from basic reflexes to more complicated behaviors like finding food and courting mates. The ability of neurons to communicate with each other as well as with other types of cells underlies all of these behaviors.

Most neurons share the same cellular components. But neurons are also highly specialized—different types of neurons have different sizes and shapes that relate to their functional roles.

Parts of a Neuron

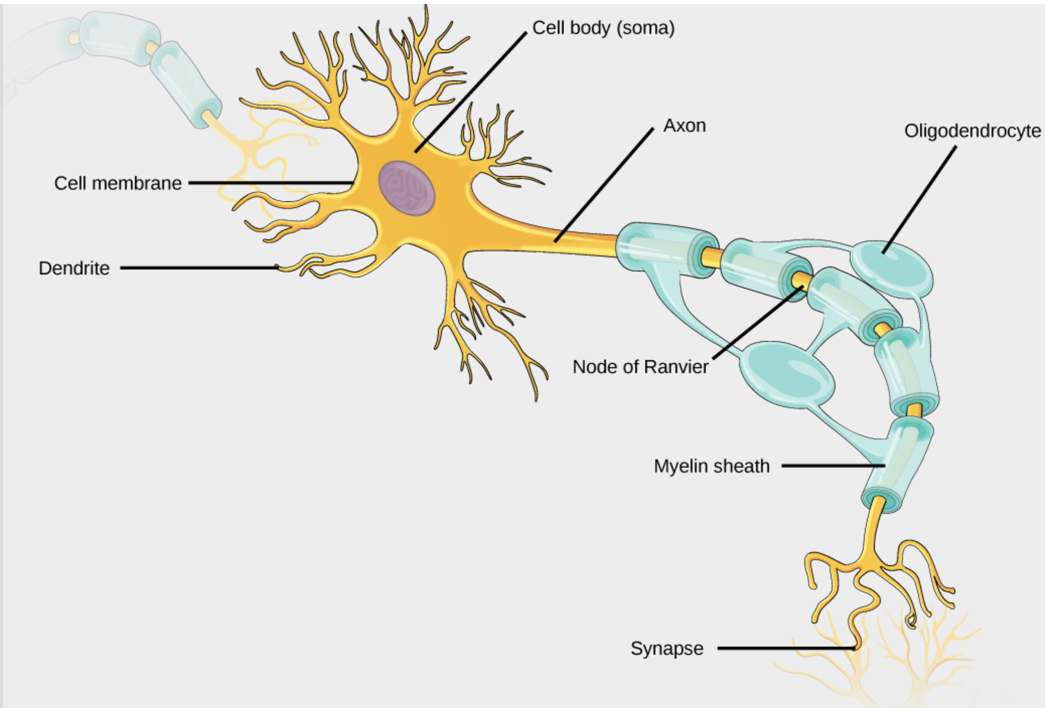
Like other cells, each neuron has a cell body (or soma) that contains a nucleus, smooth and rough endoplasmic reticulum, Golgi apparatus, mitochondria, and other cellular components. Neurons also contain unique structures, illustrated in [link] for receiving and sending the electrical signals that make neuronal communication possible. **Dendrites** are tree-like structures that extend away from the cell body to receive messages from other neurons at specialized junctions called **synapses**. Although some neurons do not have any dendrites, some types of neurons have multiple

dendrites. Dendrites can have small protrusions called dendritic spines, which further increase surface area for possible synaptic connections.

Once a signal is received by the dendrite, it then travels passively to the cell body. The cell body contains a specialized structure, the **axon hillock** that integrates signals from multiple synapses and serves as a junction between the cell body and an **axon**. An axon is a tube-like structure that propagates the integrated signal to specialized endings called **axon terminals**. These terminals in turn synapse on other neurons, muscle, or target organs. Chemicals released at axon terminals allow signals to be communicated to these other cells. Neurons usually have one or two axons, but some neurons, like amacrine cells in the retina, do not contain any axons. Some axons are covered with **myelin**, which acts as an insulator to minimize dissipation of the electrical signal as it travels down the axon, greatly increasing the speed on conduction. This insulation is important as the axon from a human motor neuron can be as long as a meter—from the base of the spine to the toes. The myelin sheath is not actually part of the neuron. Myelin is produced by glial cells. Along the axon there are periodic gaps in the myelin sheath. These gaps are called **nodes of Ranvier** and are sites where the signal is "recharged" as it travels along the axon.

It is important to note that a single neuron does not act alone—neuronal communication depends on the connections that neurons make with one another (as well as with other cells, like muscle cells). Dendrites from a single neuron may receive synaptic contact from many other neurons. For example, dendrites from a Purkinje cell in the cerebellum are thought to receive contact from as many as 200,000 other neurons.

Note:
Art Connection



Neurons contain organelles common to many other cells, such as a nucleus and mitochondria. They also have more specialized structures, including dendrites and axons.

Which of the following statements is false?

a. The soma is the cell body of a nerve cell.
b. Myelin sheath provides an insulating layer to the dendrites.
c. Axons carry the signal from the soma to the target.
d. Dendrites carry the signal to the soma.

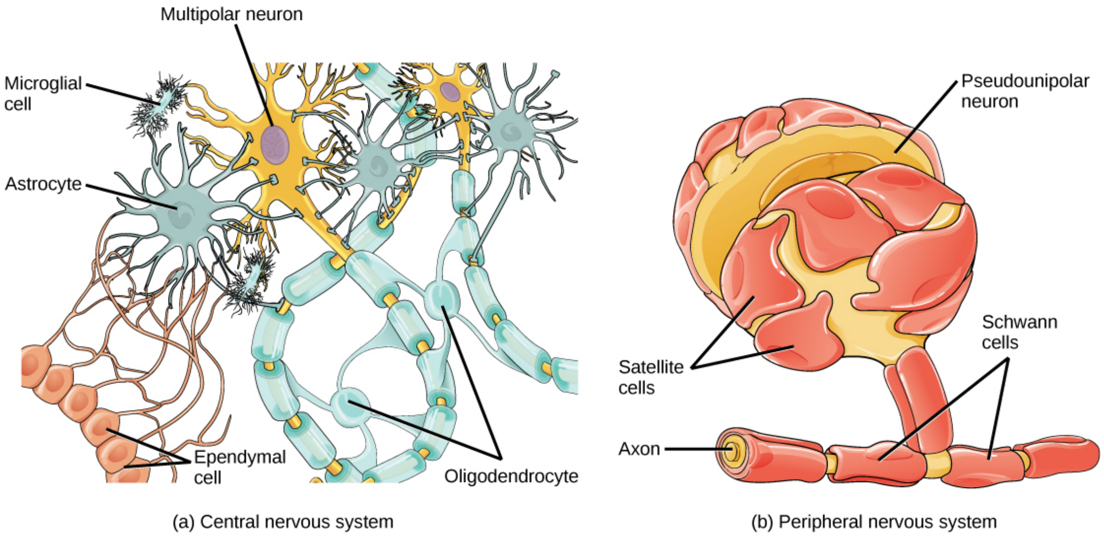
Glia

While glia are often thought of as the supporting cast of the nervous system, the number of glial cells in the brain actually outnumbers the number of neurons by a factor of ten. Neurons would be unable to function without the vital roles that are fulfilled by these glial cells. Glia guide developing

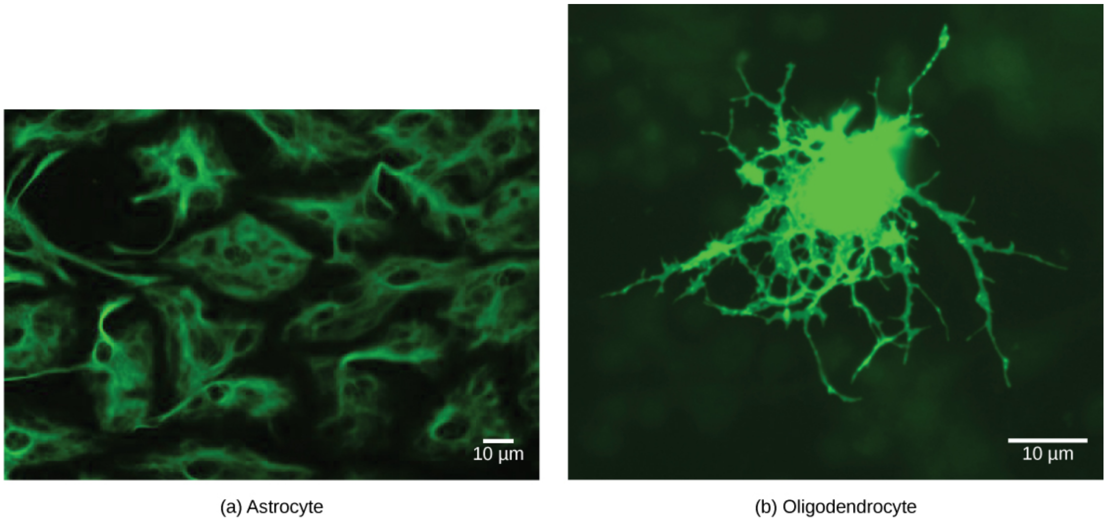
neurons to their destinations, buffer ions and chemicals that would otherwise harm neurons, and provide myelin sheaths around axons. Scientists have recently discovered that they also play a role in responding to nerve activity and modulating communication between nerve cells. When glia do not function properly, the result can be disastrous—most brain tumors are caused by mutations in glia.

Types of Glia

There are several different types of glia with different functions, two of which are shown in [link]. **Astrocytes**, shown in [link]**a** make contact with both capillaries and neurons in the CNS. They provide nutrients and other substances to neurons, regulate the concentrations of ions and chemicals in the extracellular fluid, and provide structural support for synapses. Astrocytes also form the blood-brain barrier—a structure that blocks entrance of toxic substances into the brain. Astrocytes, in particular, have been shown through calcium imaging experiments to become active in response to nerve activity, transmit calcium waves between astrocytes, and modulate the activity of surrounding synapses. **Satellite glia** provide nutrients and structural support for neurons in the PNS. **Microglia** scavenge and degrade dead cells and protect the brain from invading microorganisms. **Oligodendrocytes**, shown in [link]**b** form myelin sheaths around axons in the CNS. One axon can be myelinated by several oligodendrocytes, and one oligodendrocyte can provide myelin for multiple neurons. This is distinctive from the PNS where a single **Schwann cell** provides myelin for only one axon as the entire Schwann cell surrounds the axon. **Radial glia** serve as scaffolds for developing neurons as they migrate to their end destinations. **Ependymal** cells line fluid-filled ventricles of the brain and the central canal of the spinal cord. They are involved in the production of cerebrospinal fluid, which serves as a cushion for the brain, moves the fluid between the spinal cord and the brain, and is a component for the choroid plexus.



Glial cells support neurons and maintain their environment. Glial cells of the (a) central nervous system include oligodendrocytes, astrocytes, ependymal cells, and microglial cells. Oligodendrocytes form the myelin sheath around axons. Astrocytes provide nutrients to neurons, maintain their extracellular environment, and provide structural support. Microglia scavenge pathogens and dead cells. Ependymal cells produce cerebrospinal fluid that cushions the neurons. Glial cells of the (b) peripheral nervous system include Schwann cells, which form the myelin sheath, and satellite cells, which provide nutrients and structural support to neurons.



(a) Astrocytes and (b) oligodendrocytes are glial cells of the central nervous system. (credit a: modification of work by Uniformed Services University; credit b: modification of work by Jurjen Broeke; scale-bar data from Matt Russell)

Section Summary

The nervous system is made up of neurons and glia. Neurons are specialized cells that are capable of sending electrical as well as chemical signals. Most neurons contain dendrites, which receive these signals, and axons that send signals to other neurons or tissues. There are four main types of neurons: unipolar, bipolar, multipolar, and pseudounipolar neurons. Glia are non-neuronal cells in the nervous system that support neuronal development and signaling. There are several types of glia that serve different functions.

Art Connections

Exercise:

Problem: [link] Which of the following statements is false?

a. The soma is the cell body of a nerve cell.
b. Myelin sheath provides an insulating layer to the dendrites.
c. Axons carry the signal from the soma to the target.
d. Dendrites carry the signal to the soma.

Solution:

[link] B

Review Questions

Exercise:

Problem:

Neurons contain _________, which can receive signals from other neurons.

a. axons
b. mitochondria
c. dendrites
d. Golgi bodies

Solution:

C

Exercise:

Problem:

A(n) _________ neuron has one axon and one dendrite extending directly from the cell body.

a. unipolar
b. bipolar
c. multipolar

d. pseudounipolar

Solution:

B

Exercise:

Problem:

Glia that provide myelin for neurons in the brain are called ________.

a. Schwann cells
b. oligodendrocytes
c. microglia
d. astrocytes

Solution:

B

Free Response

Exercise:

Problem:

How are neurons similar to other cells? How are they unique?

Solution:

Neurons contain organelles common to all cells, such as a nucleus and mitochondria. They are unique because they contain dendrites, which can receive signals from other neurons, and axons that can send these signals to other cells.

Exercise:

Problem:

Multiple sclerosis causes demyelination of axons in the brain and spinal cord. Why is this problematic?

Solution:

Myelin provides insulation for signals traveling along axons. Without myelin, signal transmission can slow down and degrade over time. This would slow down neuronal communication across the nervous system and affect all downstream functions.

Glossary

astrocyte
: glial cell in the central nervous system that provide nutrients, extracellular buffering, and structural support for neurons; also makes up the blood-brain barrier

axon
: tube-like structure that propagates a signal from a neuron’s cell body to axon terminals

axon hillock
: electrically sensitive structure on the cell body of a neuron that integrates signals from multiple neuronal connections

axon terminal
: structure on the end of an axon that can form a synapse with another neuron

dendrite
: structure that extends away from the cell body to receive messages from other neurons

ependymal

cell that lines fluid-filled ventricles of the brain and the central canal of the spinal cord; involved in production of cerebrospinal fluid

glia
: (also, glial cells) cells that provide support functions for neurons

microglia
: glia that scavenge and degrade dead cells and protect the brain from invading microorganisms

myelin
: fatty substance produced by glia that insulates axons

neuron
: specialized cell that can receive and transmit electrical and chemical signals

nodes of Ranvier
: gaps in the myelin sheath where the signal is recharged

oligodendrocyte
: glial cell that myelinates central nervous system neuron axons

radial glia
: glia that serve as scaffolds for developing neurons as they migrate to their final destinations

satellite glia
: glial cell that provides nutrients and structural support for neurons in the peripheral nervous system

Schwann cell
: glial cell that creates myelin sheath around a peripheral nervous system neuron axon

synapse
: junction between two neurons where neuronal signals are communicated

Derived copy of How Neurons Communicate
By the end of this section, you will be able to:

- Describe the basis of the resting membrane potential
- Explain the stages of an action potential and how action potentials are propagated
- Explain the similarities and differences between chemical and electrical synapses
- Describe long-term potentiation and long-term depression

All functions performed by the nervous system—from a simple motor reflex to more advanced functions like making a memory or a decision—require neurons to communicate with one another. While humans use words and body language to communicate, neurons use electrical and chemical signals. Just like a person in a committee, one neuron usually receives and synthesizes messages from multiple other neurons before “making the decision” to send the message on to other neurons.

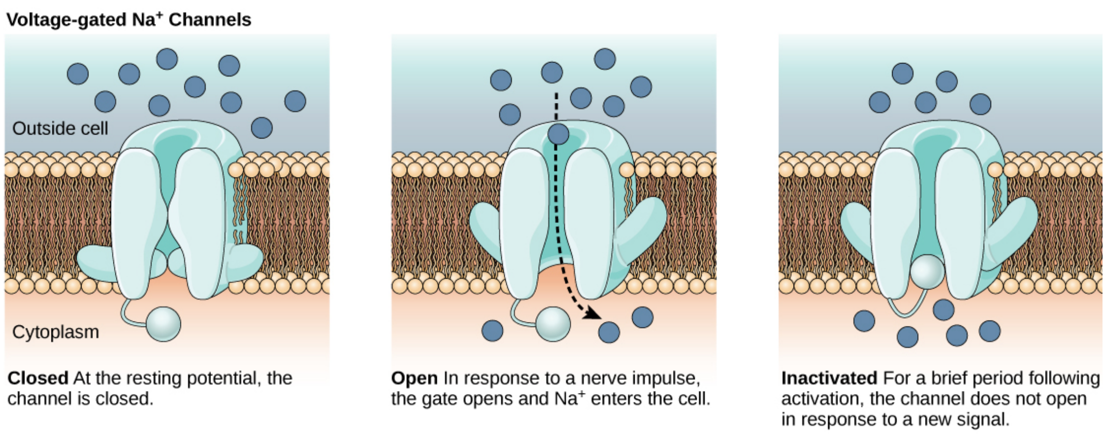
Nerve Impulse Transmission within a Neuron

For the nervous system to function, neurons must be able to send and receive signals. These signals are possible because each neuron has a charged cellular membrane (a voltage difference between the inside and the outside), and the charge of this membrane can change in response to neurotransmitter molecules released from other neurons and environmental stimuli. To understand how neurons communicate, one must first understand the basis of the baseline or ‘resting’ membrane charge.

Neuronal Charged Membranes

The lipid bilayer membrane that surrounds a neuron is impermeable to charged molecules or ions. To enter or exit the neuron, ions must pass through special proteins called ion channels that span the membrane. Ion channels have different configurations: open, closed, and inactive, as illustrated in [link]. Some ion channels need to be activated in order to open and allow ions to pass into or out of the cell. These ion channels are

sensitive to the environment and can change their shape accordingly. Ion channels that change their structure in response to voltage changes are called voltage-gated ion channels. Voltage-gated ion channels regulate the relative concentrations of different ions inside and outside the cell. The difference in total charge between the inside and outside of the cell is called the **membrane potential**.



Voltage-gated ion channels open in response to changes in membrane voltage. After activation, they become inactivated for a brief period and will no longer open in response to a signal.

Note:
Link to Learning

This video discusses the basis of the resting membrane potential.

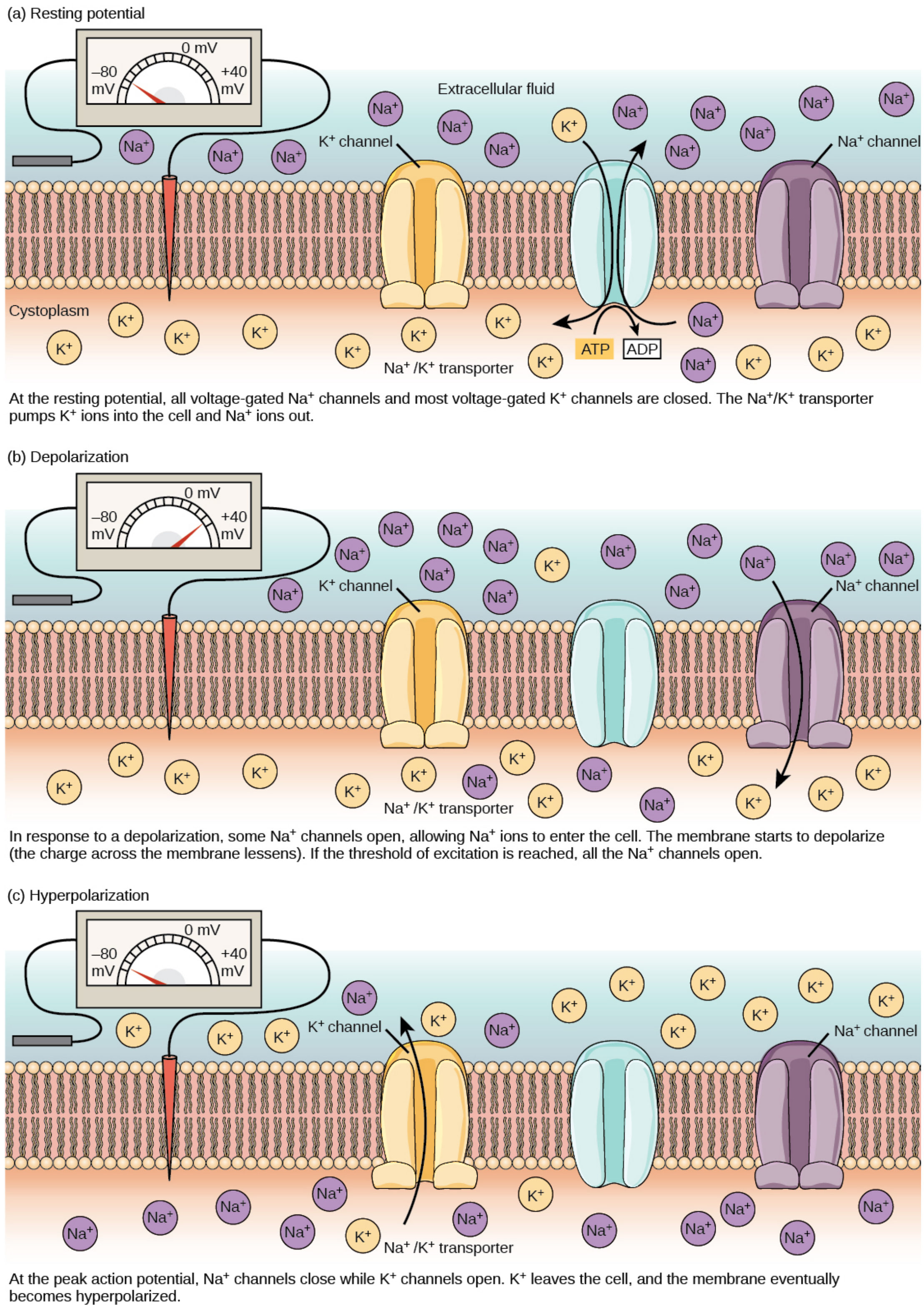
Resting Membrane Potential

A neuron at rest is negatively charged: the inside of a cell is approximately 70 millivolts more negative than the outside (−70 mV, note that this number varies by neuron type and by species). This voltage is called the resting membrane potential; it is caused by differences in the concentrations of ions inside and outside the cell. If the membrane were equally permeable to all ions, each type of ion would flow across the membrane and the system would reach equilibrium. Because ions cannot simply cross the membrane at will, there are different concentrations of several ions inside and outside the cell, as shown in [link]. The difference in the number of positively charged potassium ions (K^+) inside and outside the cell dominates the resting membrane potential ([link]). When the membrane is at rest, K^+ ions accumulate inside the cell due to a net movement with the concentration gradient. The negative resting membrane potential is created and maintained by increasing the concentration of cations outside the cell (in the extracellular fluid) relative to inside the cell (in the cytoplasm). The negative charge within the cell is created by the cell membrane being more permeable to potassium ion movement than sodium ion movement. In neurons, potassium ions are maintained at high concentrations within the cell while sodium ions are maintained at high concentrations outside of the cell. The cell possesses potassium and sodium leakage channels that allow the two cations to diffuse down their concentration gradient. However, the neurons have far more potassium leakage channels than sodium leakage channels. Therefore, potassium diffuses out of the cell at a much faster rate than sodium leaks in. Because more cations are leaving the cell than are entering, this causes the interior of the cell to be negatively charged relative to the outside of the cell. The actions of the sodium potassium pump help to maintain the resting potential, once established. Recall that sodium potassium pumps brings two K^+ ions into the cell while removing three Na^+ ions per ATP consumed. As more cations are expelled from the cell than taken in, the inside of the cell remains negatively charged relative to the extracellular fluid. It should be noted that calcium ions (Cl^-) tend to

accumulate outside of the cell because they are repelled by negatively-charged proteins within the cytoplasm.

Ion Concentration Inside and Outside Neurons			
Ion	**Extracellular concentration (mM)**	**Intracellular concentration (mM)**	**Ratio outside/inside**
Na^+	145	12	12
K+	4	155	0.026
Cl^-	120	4	30
Organic anions (A−)	—	100	

The resting membrane potential is a result of different concentrations inside and outside the cell.



The (a) resting membrane potential is a result of different concentrations of Na^+ and K^+ ions inside and outside the cell.

A nerve impulse causes Na^+ to enter the cell, resulting in (b) depolarization. At the peak action potential, K^+ channels open and the cell becomes (c) hyperpolarized.

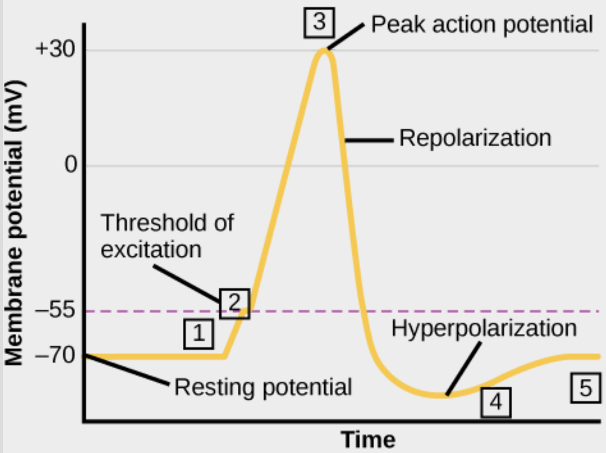
Action Potential

A neuron can receive input from other neurons and, if this input is strong enough, send the signal to downstream neurons. Transmission of a signal between neurons is generally carried by a chemical called a neurotransmitter. Transmission of a signal within a neuron (from dendrite to axon terminal) is carried by a brief reversal of the resting membrane potential called an **action potential**. When neurotransmitter molecules bind to receptors located on a neuron's dendrites, ion channels open. At excitatory synapses, this opening allows positive ions to enter the neuron and results in **depolarization** of the membrane—a decrease in the difference in voltage between the inside and outside of the neuron. A stimulus from a sensory cell or another neuron depolarizes the target neuron to its threshold potential (-55 mV). Na^+ channels in the axon hillock open, allowing positive ions to enter the cell ([link] and [link]). Once the sodium channels open, the neuron completely depolarizes to a membrane potential of about +40 mV. Action potentials are considered an "all-or nothing" event, in that, once the threshold potential is reached, the neuron always completely depolarizes. Once depolarization is complete, the cell must now "reset" its membrane voltage back to the resting potential. To accomplish this, the Na^+ channels close and cannot be opened. This begins the neuron's **refractory period**, in which it cannot produce another action potential because its sodium channels will not open. At the same time, voltage-gated K^+ channels open, allowing K^+ to leave the cell. As K^+ ions leave the cell, the membrane potential once again becomes negative. The diffusion of K^+ out of the cell actually **hyperpolarizes** the cell, in that the membrane potential becomes more negative than the cell's normal resting potential. At this point, the sodium channels will return to their resting state, meaning they are ready to open again if the membrane potential again exceeds the

threshold potential. Eventually the extra K^+ ions diffuse out of the cell through the potassium leakage channels, bringing the cell from its hyperpolarized state, back to its resting membrane potential.

Note:

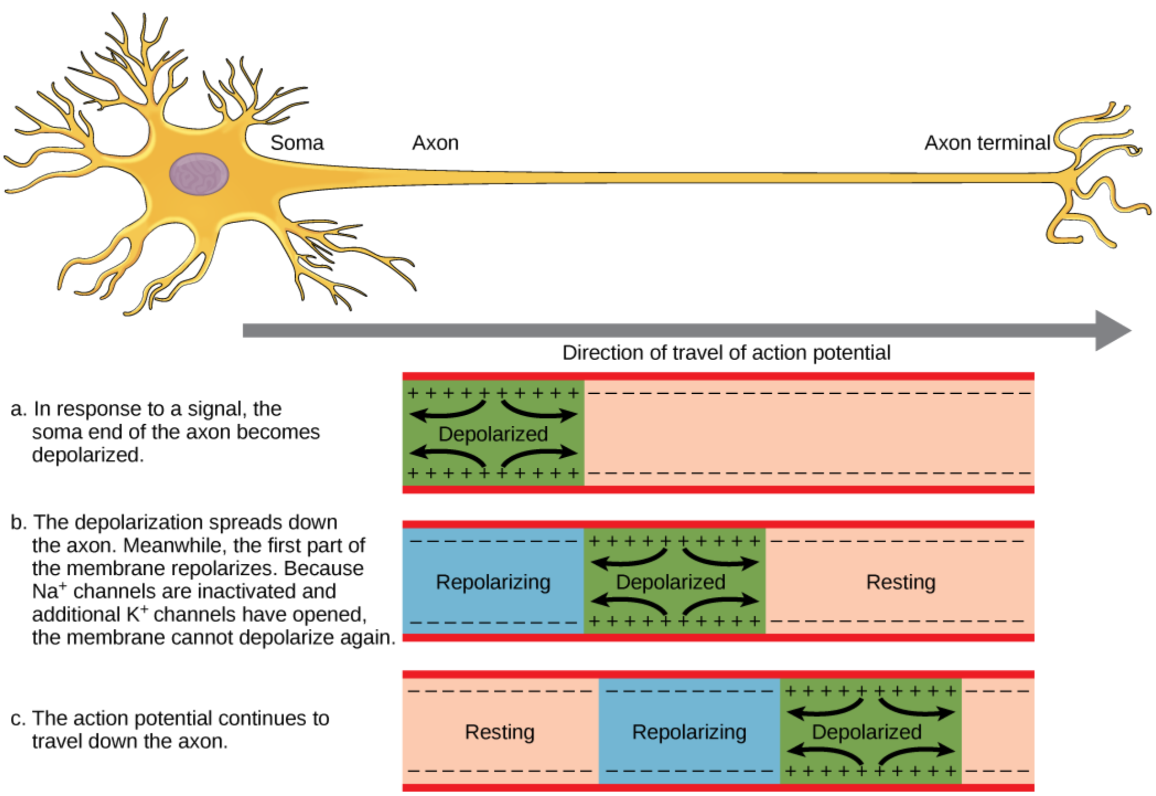
Art Connection



The formation of an action potential can be divided into five steps: (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all Na^+ channels open and the membrane depolarizes. (3) At the peak action potential, K^+ channels open and K^+ begins to leave the cell. At the same time, Na^+ channels close. (4) The membrane becomes hyperpolarized as K^+ ions continue to leave the cell. The hyperpolarized membrane is in

a refractory period and cannot fire. (5) The K^+ channels close and the Na^+/K^+ transporter restores the resting potential.

Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of K^+ through voltage-gated K^+ channels. Which part of the action potential would you expect potassium channels to affect?



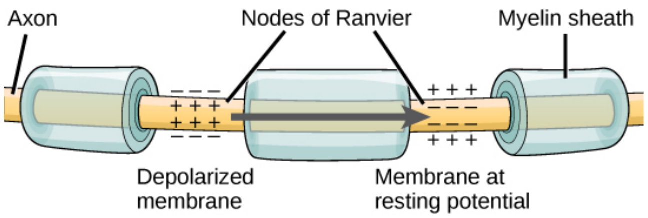
The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes.

Note:
Link to Learning

This video presents an overview of action potential.

Myelin and the Propagation of the Action Potential

For an action potential to communicate information to another neuron, it must travel along the axon and reach the axon terminals where it can initiate neurotransmitter release. The speed of conduction of an action potential along an axon is influenced by both the diameter of the axon and the axon’s resistance to current leak. Myelin acts as an insulator that prevents current from leaving the axon; this increases the speed of action potential conduction. In demyelinating diseases like multiple sclerosis, action potential conduction slows because current leaks from previously insulated axon areas. The nodes of Ranvier, illustrated in [link] are gaps in the myelin sheath along the axon. These unmyelinated spaces are about one micrometer long and contain voltage gated Na^+ and K^+ channels. Flow of ions through these channels, particularly the Na^+ channels, regenerates the action potential over and over again along the axon. This ‘jumping’ of the action potential from one node to the next is called **saltatory conduction**. If nodes of Ranvier were not present along an axon, the action potential would propagate very slowly since Na^+ and K^+ channels would have to continuously regenerate action potentials at every point along the axon instead of at specific points. Nodes of Ranvier also save energy for the neuron since the channels only need to be present at the nodes and not along the entire axon.



Nodes of Ranvier are gaps in myelin coverage along axons. Nodes contain voltage-gated K^+ and Na^+ channels. Action potentials travel down the axon by jumping from one node to the next.

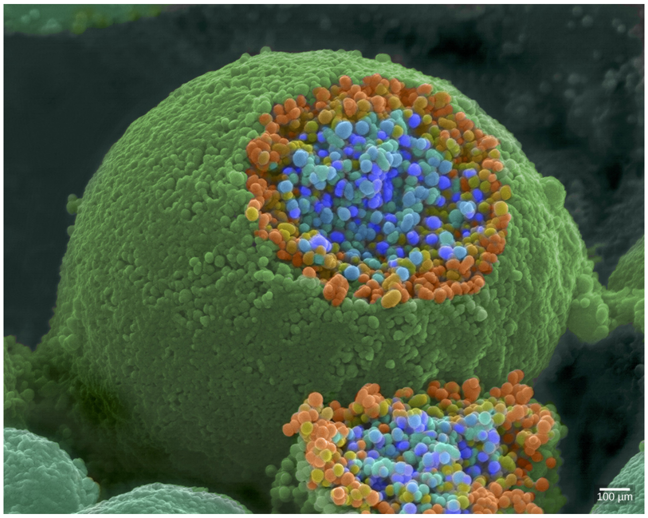
Synaptic Transmission

The synapse or "gap" is the place where information is transmitted from one neuron to another. Synapses usually form between axon terminals and dendritic spines, but this is not universally true. There are also axon-to-axon, dendrite-to-dendrite, and axon-to-cell body synapses. The neuron transmitting the signal is called the presynaptic neuron, and the neuron receiving the signal is called the postsynaptic neuron. Note that these designations are relative to a particular synapse—most neurons are both presynaptic and postsynaptic. There are two types of synapses: chemical and electrical.

Chemical Synapse

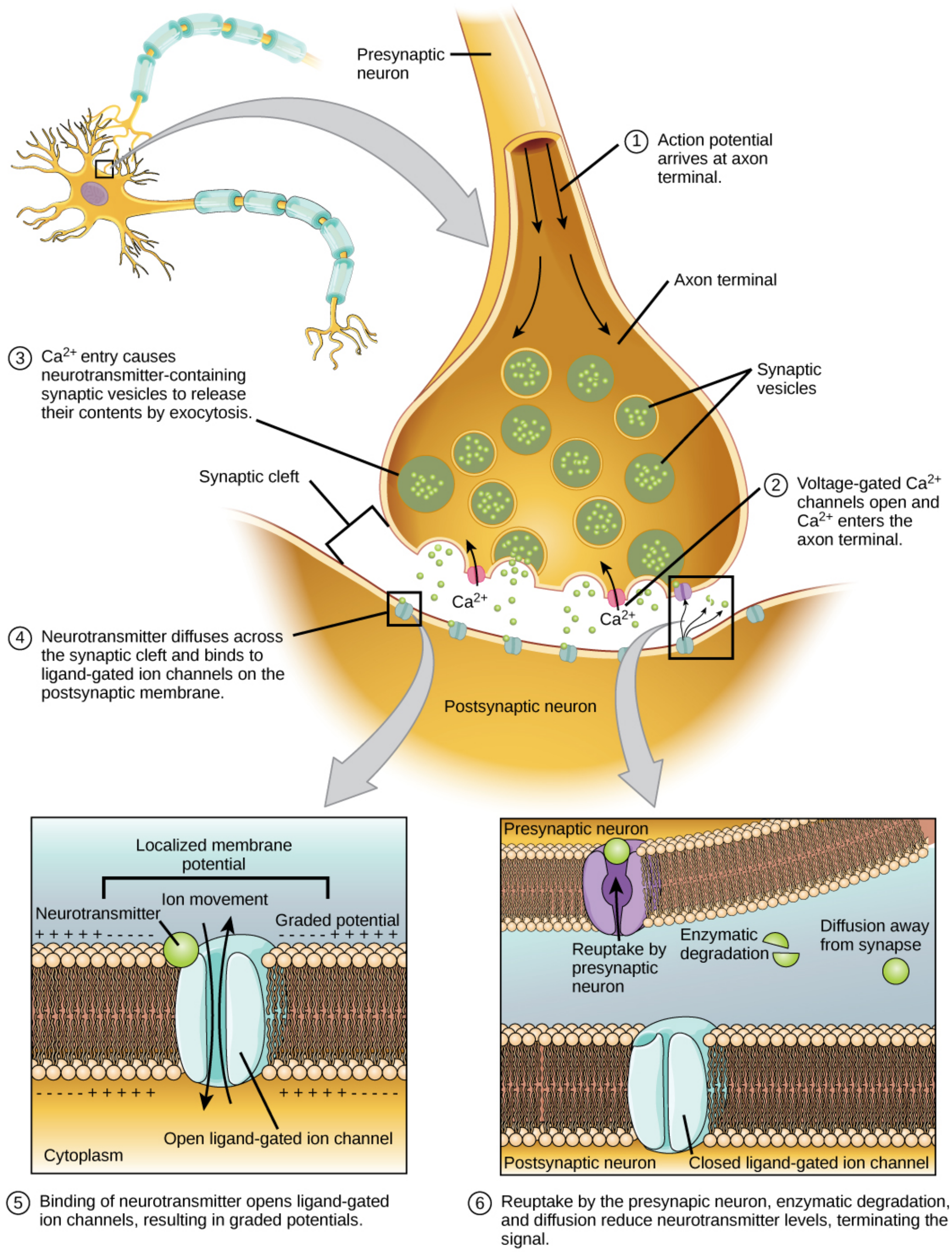
When an action potential reaches the axon terminal it depolarizes the membrane and opens voltage-gated Na^+ channels. Na^+ ions enter the cell, further depolarizing the presynaptic membrane. This depolarization causes voltage-gated Ca^{2+} channels to open. Calcium ions entering the cell initiate a signaling cascade that causes small membrane-bound vesicles, called

synaptic vesicles, containing neurotransmitter molecules to fuse with the presynaptic membrane. Synaptic vesicles are shown in [link], which is an image from a scanning electron microscope.



This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (credit: modification of work by Tina Carvalho, NIH-NIGMS; scale-bar data from Matt Russell)

Fusion of a vesicle with the presynaptic membrane causes neurotransmitter to be released into the **synaptic cleft**, the extracellular space between the presynaptic and postsynaptic membranes, as illustrated in [link]. The neurotransmitter diffuses across the synaptic cleft and binds to receptor proteins on the postsynaptic membrane.



Communication at chemical synapses requires release of neurotransmitters. When the presynaptic membrane is depolarized, voltage-gated Ca^{2+} channels open and allow Ca^{2+} to enter the cell. The calcium entry causes synaptic vesicles to fuse

with the membrane and release neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the synaptic cleft and binds to ligand-gated ion channels in the postsynaptic membrane, resulting in a localized depolarization or hyperpolarization of the postsynaptic neuron.

The binding of a specific neurotransmitter causes particular ion channels, in this case ligand-gated channels, on the postsynaptic membrane to open. Neurotransmitters can either have excitatory or inhibitory effects on the postsynaptic membrane, as detailed in [link]. For example, when acetylcholine is released at the synapse between a nerve and muscle (called the neuromuscular junction) by a presynaptic neuron, it causes postsynaptic Na^+ channels to open. Na^+ enters the postsynaptic cell and causes the postsynaptic membrane to depolarize. This depolarization is called an **excitatory postsynaptic potential (EPSP)** and makes the postsynaptic neuron more likely to fire an action potential. Release of neurotransmitter at inhibitory synapses causes **inhibitory postsynaptic potentials (IPSPs)**, a hyperpolarization of the presynaptic membrane. For example, when the neurotransmitter GABA (gamma-aminobutyric acid) is released from a presynaptic neuron, it binds to and opens Cl^- channels. Cl^- ions enter the cell and hyperpolarizes the membrane, making the neuron less likely to fire an action potential.

Once neurotransmission has occurred, the neurotransmitter must be removed from the synaptic cleft so the postsynaptic membrane can “reset” and be ready to receive another signal. This can be accomplished in three ways: the neurotransmitter can diffuse away from the synaptic cleft, it can be degraded by enzymes in the synaptic cleft, or it can be recycled (sometimes called reuptake) by the presynaptic neuron. Several drugs act at this step of neurotransmission. For example, some drugs that are given to Alzheimer’s patients work by inhibiting acetylcholinesterase, the enzyme that degrades acetylcholine. This inhibition of the enzyme essentially increases neurotransmission at synapses that release acetylcholine. Once released, the acetylcholine stays in the cleft and can continually bind and unbind to postsynaptic receptors.

Neurotransmitter Function and Location		
Neurotransmitter	**Example**	**Location**
Acetylcholine	—	CNS and/or PNS
Biogenic amine	Dopamine, serotonin, norepinephrine	CNS and/or PNS
Amino acid	Glycine, glutamate, aspartate, gamma aminobutyric acid	CNS
Neuropeptide	Substance P, endorphins	CNS and/or PNS

Electrical Synapse

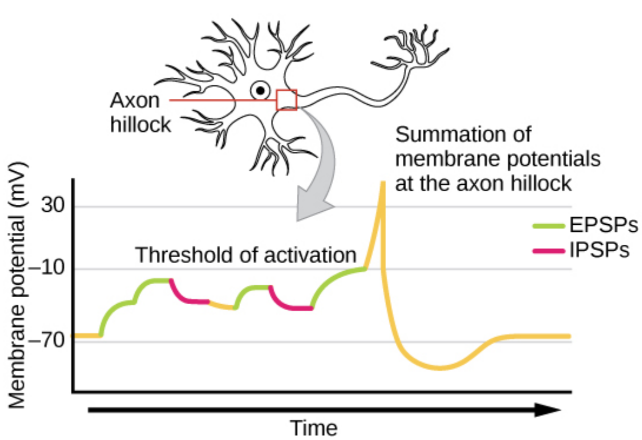
While electrical synapses are fewer in number than chemical synapses, they are found in all nervous systems and play important and unique roles. The mode of neurotransmission in electrical synapses is quite different from that in chemical synapses. In an electrical synapse, the presynaptic and postsynaptic membranes are very close together and are actually physically connected by channel proteins forming gap junctions. Gap junctions allow current to pass directly from one cell to the next. In addition to the ions that carry this current, other molecules, such as ATP, can diffuse through the large gap junction pores.

There are key differences between chemical and electrical synapses. Because chemical synapses depend on the release of neurotransmitter molecules from synaptic vesicles to pass on their signal, there is an approximately one millisecond delay between when the axon potential

reaches the presynaptic terminal and when the neurotransmitter leads to opening of postsynaptic ion channels. Additionally, this signaling is unidirectional. Signaling in electrical synapses, in contrast, is virtually instantaneous (which is important for synapses involved in key reflexes), and some electrical synapses are bidirectional. Electrical synapses are also more reliable as they are less likely to be blocked, and they are important for synchronizing the electrical activity of a group of neurons. For example, electrical synapses in the thalamus are thought to regulate slow-wave sleep, and disruption of these synapses can cause seizures.

Signal Summation

Sometimes a single EPSP is strong enough to induce an action potential in the postsynaptic neuron, but often multiple presynaptic inputs must create EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential. This process is called **summation** and occurs at the axon hillock, as illustrated in [link]. Additionally, one neuron often has inputs from many presynaptic neurons—some excitatory and some inhibitory—so IPSPs can cancel out EPSPs and vice versa. It is the net change in postsynaptic membrane voltage that determines whether the postsynaptic cell has reached its threshold of excitation needed to fire an action potential. Together, synaptic summation and the threshold for excitation act as a filter so that random “noise” in the system is not transmitted as important information.

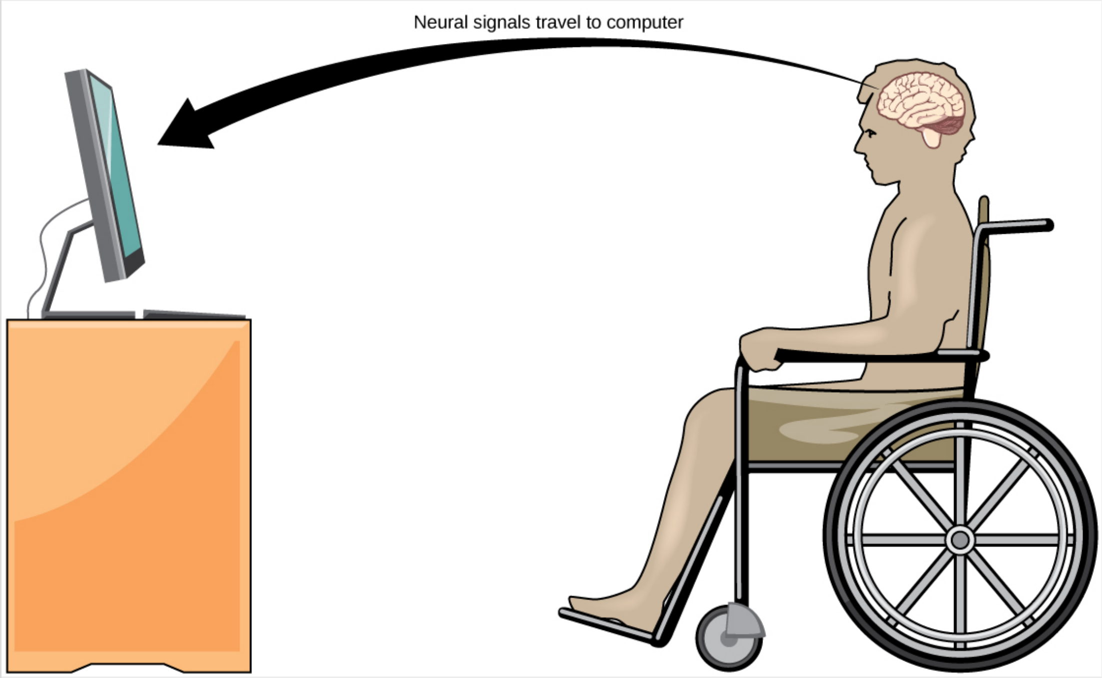


A single neuron can receive both excitatory and inhibitory inputs from multiple neurons, resulting in local membrane depolarization (EPSP input) and hyperpolarization (IPSP input). All these inputs are added together at the axon hillock. If the EPSPs are strong enough to overcome the IPSPs and reach the threshold of excitation, the neuron will fire.

Note:
Everyday Connection
Brain-computer interface
Amyotrophic lateral sclerosis (ALS, also called Lou Gehrig’s Disease) is a neurological disease characterized by the degeneration of the motor neurons that control voluntary movements. The disease begins with muscle weakening and lack of coordination and eventually destroys the neurons that control speech, breathing, and swallowing; in the end, the disease can lead to paralysis. At that point, patients require assistance from machines to be able to breathe and to communicate. Several special technologies have been developed to allow “locked-in” patients to communicate with the rest of the world. One technology, for example, allows patients to type out sentences by twitching their cheek. These sentences can then be read aloud by a computer.
A relatively new line of research for helping paralyzed patients, including those with ALS, to communicate and retain a degree of self-sufficiency is called brain-computer interface (BCI) technology and is illustrated in [link]. This technology sounds like something out of science fiction: it allows paralyzed patients to control a computer using only their thoughts. There are several forms of BCI. Some forms use EEG recordings from electrodes taped onto the skull. These recordings contain information from

large populations of neurons that can be decoded by a computer. Other forms of BCI require the implantation of an array of electrodes smaller than a postage stamp in the arm and hand area of the motor cortex. This form of BCI, while more invasive, is very powerful as each electrode can record actual action potentials from one or more neurons. These signals are then sent to a computer, which has been trained to decode the signal and feed it to a tool—such as a cursor on a computer screen. This means that a patient with ALS can use e-mail, read the Internet, and communicate with others by thinking of moving his or her hand or arm (even though the paralyzed patient cannot make that bodily movement). Recent advances have allowed a paralyzed locked-in patient who suffered a stroke 15 years ago to control a robotic arm and even to feed herself coffee using BCI technology.

Despite the amazing advancements in BCI technology, it also has limitations. The technology can require many hours of training and long periods of intense concentration for the patient; it can also require brain surgery to implant the devices.



With brain-computer interface technology, neural signals from a paralyzed patient are collected, decoded, and then fed to a tool, such as a computer, a wheelchair, or a robotic arm.

Note:
Link to Learning

Watch this video in which a paralyzed woman use a brain-controlled robotic arm to bring a drink to her mouth, among other images of brain-computer interface technology in action.

Art Connections

Exercise:

Problem:

[link] Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of K+ through voltage-gated K+ channels. Which part of the action potential would you expect potassium channels to affect?

Solution:

[link] Potassium channel blockers slow the repolarization phase, but have no effect on depolarization.

Review Questions

Exercise:

Problem:

For a neuron to fire an action potential, its membrane must reach _________.

a. hyperpolarization
b. the threshold of excitation
c. the refractory period
d. inhibitory postsynaptic potential

Solution:

B

Exercise:

Problem:

After an action potential, the opening of additional voltage-gated _________ channels and the inactivation of sodium channels, cause the membrane to return to its resting membrane potential.

a. sodium
b. potassium
c. calcium
d. chloride

Solution:

B

Exercise:

Problem:

What is the term for protein channels that connect two neurons at an electrical synapse?

a. synaptic vesicles
b. voltage-gated ion channels
c. gap junction protein
d. sodium-potassium exchange pumps

Solution:

C

Free Response

Exercise:

Problem:

How does myelin aid propagation of an action potential along an axon? How do the nodes of Ranvier help this process?

Solution:

Myelin prevents the leak of current from the axon. Nodes of Ranvier allow the action potential to be regenerated at specific points along the axon. They also save energy for the cell since voltage-gated ion channels and sodium-potassium transporters are not needed along myelinated portions of the axon.

Exercise:

Problem: What are the main steps in chemical neurotransmission?

Solution:

An action potential travels along an axon until it depolarizes the membrane at an axon terminal. Depolarization of the membrane causes voltage-gated Ca^{2+} channels to open and Ca^{2+} to enter the cell. The intracellular calcium influx causes synaptic vesicles containing neurotransmitter to fuse with the presynaptic membrane. The

neurotransmitter diffuses across the synaptic cleft and binds to receptors on the postsynaptic membrane. Depending on the specific neurotransmitter and postsynaptic receptor, this action can cause positive (excitatory postsynaptic potential) or negative (inhibitory postsynaptic potential) ions to enter the cell.

Glossary

action potential
: self-propagating momentary change in the electrical potential of a neuron (or muscle) membrane

depolarization
: change in the membrane potential to a less negative value

excitatory postsynaptic potential (EPSP)
: depolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

hyperpolarization
: change in the membrane potential to a more negative value

inhibitory postsynaptic potential (IPSP)
: hyperpolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

long-term depression (LTD)
: prolonged decrease in synaptic coupling between a pre- and postsynaptic cell

long-term potentiation (LTP)
: prolonged increase in synaptic coupling between a pre-and postsynaptic cell

membrane potential
: difference in electrical potential between the inside and outside of a cell

refractory period
period after an action potential when it is more difficult or impossible for an action potential to be fired; caused by inactivation of sodium channels and activation of additional potassium channels of the membrane

saltatory conduction
“jumping” of an action potential along an axon from one node of Ranvier to the next

summation
process of multiple presynaptic inputs creating EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential

synaptic cleft
space between the presynaptic and postsynaptic membranes

synaptic vesicle
spherical structure that contains a neurotransmitter

threshold of excitation
level of depolarization needed for an action potential to fire

The Central Nervous System

By the end of this section, you will be able to:

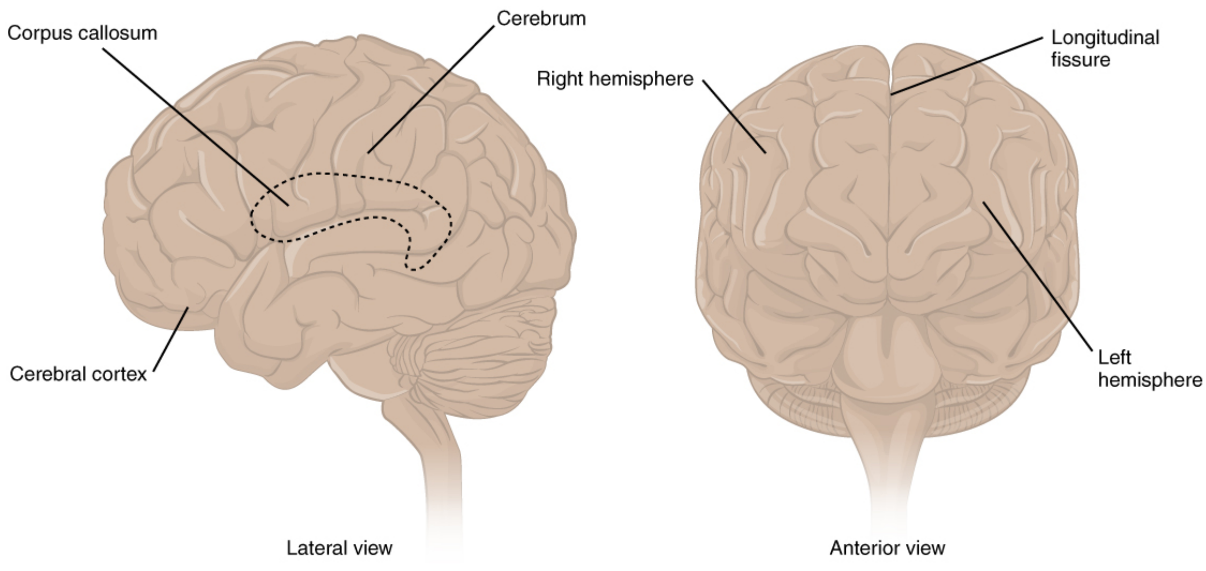
- Name the major regions of the adult brain
- Describe the connections between the cerebrum and brain stem through the diencephalon, and from those regions into the spinal cord
- Recognize the complex connections within the subcortical structures of the basal nuclei
- Explain the arrangement of gray and white matter in the spinal cord

The brain and the spinal cord are the central nervous system, and they represent the main organs of the nervous system. The spinal cord is a single structure, whereas the adult brain is described in terms of four major regions: the cerebrum, the diencephalon, the brain stem, and the cerebellum. A person's conscious experiences are based on neural activity in the brain. The regulation of homeostasis is governed by a specialized region in the brain. The coordination of reflexes depends on the integration of sensory and motor pathways in the spinal cord.

The Cerebrum

The iconic gray mantle of the human brain, which appears to make up most of the mass of the brain, is the **cerebrum** ([link]). The wrinkled portion is the **cerebral cortex**, and the rest of the structure is beneath that outer covering. There is a large separation between the two sides of the cerebrum called the **longitudinal fissure**. It separates the cerebrum into two distinct halves, a right and left **cerebral hemisphere**. Deep within the cerebrum, the white matter of the **corpus callosum** provides the major pathway for communication between the two hemispheres of the cerebral cortex.

The Cerebrum



The cerebrum is a large component of the CNS in humans, and the most obvious aspect of it is the folded surface called the cerebral cortex.

Many of the higher neurological functions, such as memory, emotion, and consciousness, are the result of cerebral function. The complexity of the cerebrum is different across vertebrate species. The cerebrum of the most primitive vertebrates is not much more than the connection for the sense of smell. In mammals, the cerebrum comprises the outer gray matter that is the cortex (from the Latin word meaning “bark of a tree”) and several deep nuclei that belong to three important functional groups. The **basal nuclei** are responsible for cognitive processing, the most important function being that associated with planning movements. The **basal forebrain** contains nuclei that are important in learning and memory. The **limbic cortex** is the region of the cerebral cortex that is part of the **limbic system**, a collection of structures involved in emotion, memory, and behavior.

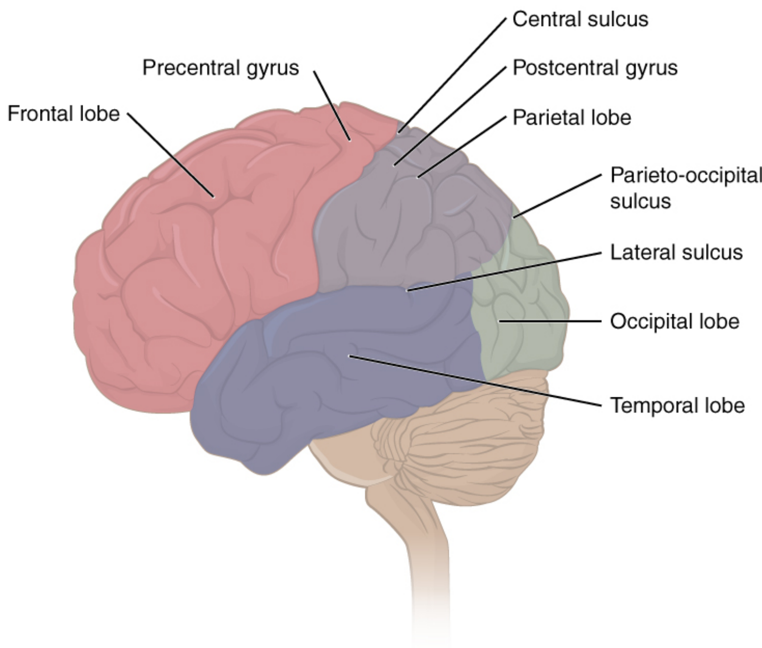
Cerebral Cortex

The cerebrum is covered by a continuous layer of gray matter that wraps around either side of the forebrain—the cerebral cortex. This thin, extensive region of wrinkled gray matter is responsible for the higher functions of the nervous system. A **gyrus** (plural = gyri) is the ridge of one of those wrinkles, and a **sulcus** (plural = sulci) is the groove between two gyri. The pattern of these folds of tissue indicates specific regions of the cerebral cortex.

The head is limited by the size of the birth canal, and the brain must fit inside the cranial cavity of the skull. Extensive folding in the cerebral cortex enables more gray matter to fit into this limited space. If the gray matter of the cortex were peeled off of the cerebrum and laid out flat, its surface area would be roughly equal to one square meter.

The folding of the cortex maximizes the amount of gray matter in the cranial cavity. During embryonic development, as the telencephalon expands within the skull, the brain goes through a regular course of growth that results in everyone's brain having a similar pattern of folds. The surface of the brain can be mapped on the basis of the locations of large gyri and sulci. Using these landmarks, the cortex can be separated into four major regions, or lobes ([link]). The **lateral sulcus** that separates the **temporal lobe** from the other regions is one such landmark. Superior to the lateral sulcus are the **parietal lobe** and **frontal lobe**, which are separated from each other by the **central sulcus**. The posterior region of the cortex is the **occipital lobe**, which has no obvious anatomical border between it and the parietal or temporal lobes on the lateral surface of the brain. From the medial surface, an obvious landmark separating the parietal and occipital lobes is called the **parieto-occipital sulcus**. The fact that there is no obvious anatomical border between these lobes is consistent with the functions of these regions being interrelated.

Lobes of the Cerebral Cortex



The cerebral cortex is divided into four lobes. Extensive folding increases the surface area available for cerebral functions.

Different regions of the cerebral cortex can be associated with particular functions, a concept known as localization of function. In the early 1900s, a German neuroscientist named Korbinian Brodmann performed an extensive study of the microscopic anatomy—the cytoarchitecture—of the cerebral cortex and divided the cortex into 52 separate regions on the basis of the histology of the cortex. His work resulted in a system of classification known as **Brodmann's areas**, which is still used today to describe the anatomical distinctions within the cortex ([link]). The results from Brodmann's work on the anatomy align very well with the functional differences within the cortex. Areas 17 and 18 in the occipital lobe are responsible for primary visual perception. That visual information is complex, so it is processed in the temporal and parietal lobes as well.

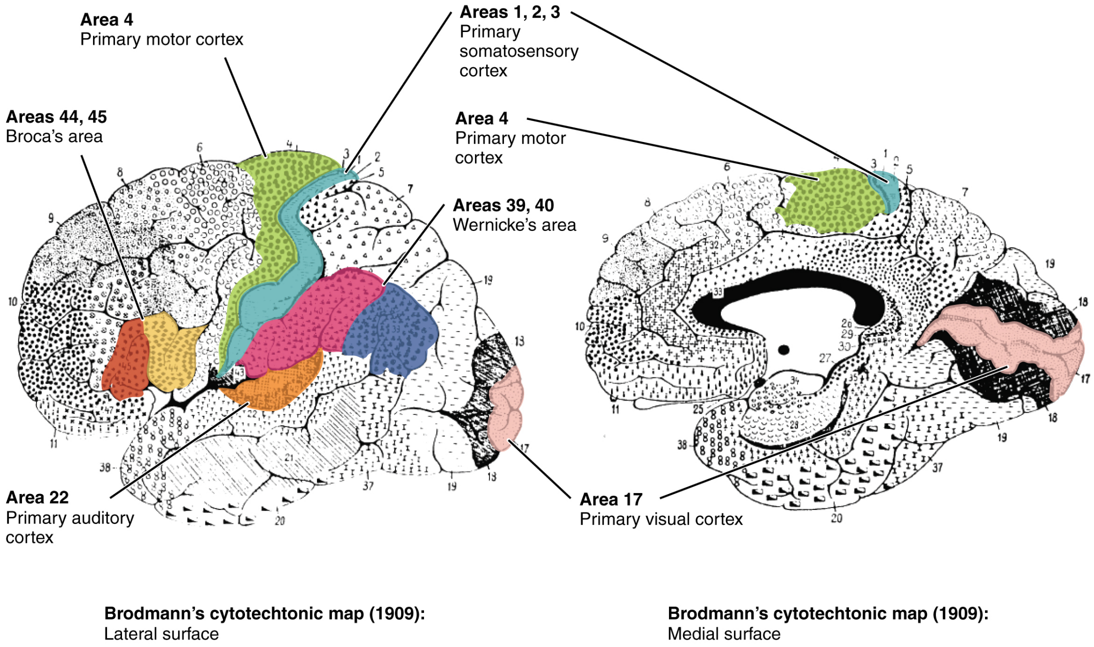
The temporal lobe is associated with primary auditory sensation, known as Brodmann's areas 41 and 42 in the superior temporal lobe. Because regions

of the temporal lobe are part of the limbic system, memory is an important function associated with that lobe. Memory is essentially a sensory function; memories are recalled sensations such as the smell of Mom’s baking or the sound of a barking dog. Even memories of movement are really the memory of sensory feedback from those movements, such as stretching muscles or the movement of the skin around a joint. Structures in the temporal lobe are responsible for establishing long-term memory, but the ultimate location of those memories is usually in the region in which the sensory perception was processed.

The main sensation associated with the parietal lobe is **somatosensation**, meaning the general sensations associated with the body. Posterior to the central sulcus is the **postcentral gyrus**, the primary somatosensory cortex, which is identified as Brodmann’s areas 1, 2, and 3. All of the tactile senses are processed in this area, including touch, pressure, tickle, pain, itch, and vibration, as well as more general senses of the body such as **proprioception** and **kinesthesia**, which are the senses of body position and movement, respectively.

Anterior to the central sulcus is the frontal lobe, which is primarily associated with motor functions. The **precentral gyrus** is the primary motor cortex. Cells from this region of the cerebral cortex are the upper motor neurons that instruct cells in the spinal cord to move skeletal muscles. Anterior to this region are a few areas that are associated with planned movements. The **premotor area** is responsible for thinking of a movement to be made. The **frontal eye fields** are important in eliciting eye movements and in attending to visual stimuli. **Broca’s area** is responsible for the production of language, or controlling movements responsible for speech; in the vast majority of people, it is located only on the left side. Anterior to these regions is the **prefrontal lobe**, which serves cognitive functions that can be the basis of personality, short-term memory, and consciousness. The prefrontal lobotomy is an outdated mode of treatment for personality disorders (psychiatric conditions) that profoundly affected the personality of the patient.

Brodmann's Areas of the Cerebral Cortex



Brodmann mapping of functionally distinct regions of the cortex was based on its cytoarchitecture at a microscopic level.

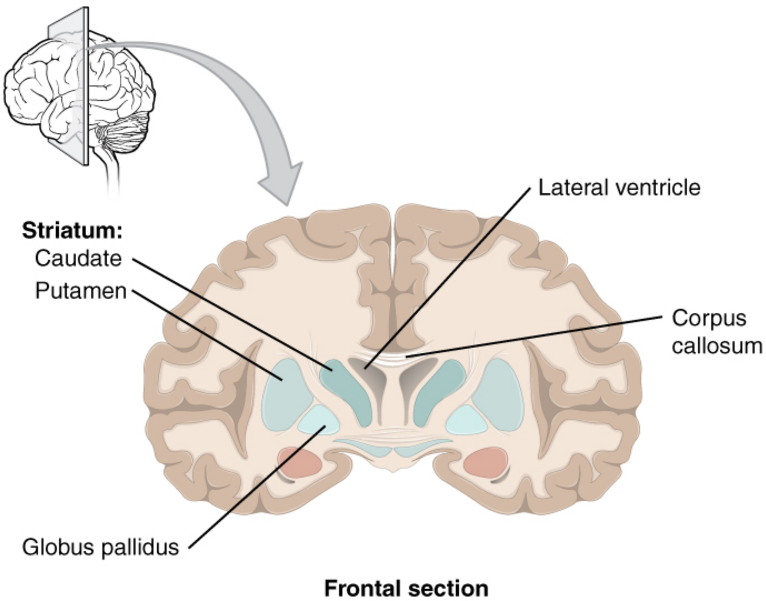
Subcortical structures

Beneath the cerebral cortex are sets of nuclei known as **subcortical nuclei** that augment cortical processes. The nuclei of the basal forebrain serve as the primary location for acetylcholine production, which modulates the overall activity of the cortex, possibly leading to greater attention to sensory stimuli. Alzheimer's disease is associated with a loss of neurons in the basal forebrain. The **hippocampus** and **amygdala** are medial-lobe structures that, along with the adjacent cortex, are involved in long-term memory formation and emotional responses. The basal nuclei are a set of nuclei in the cerebrum responsible for comparing cortical processing with the general state of activity in the nervous system to influence the likelihood of movement taking place. For example, while a student is sitting in a classroom listening to a lecture, the basal nuclei will keep the urge to jump up and scream from actually happening. (The basal nuclei are also referred

to as the basal ganglia, although that is potentially confusing because the term ganglia is typically used for peripheral structures.)

The major structures of the basal nuclei that control movement are the **caudate**, **putamen**, and **globus pallidus**, which are located deep in the cerebrum. The caudate is a long nucleus that follows the basic C-shape of the cerebrum from the frontal lobe, through the parietal and occipital lobes, into the temporal lobe. The putamen is mostly deep in the anterior regions of the frontal and parietal lobes. Together, the caudate and putamen are called the **striatum**. The globus pallidus is a layered nucleus that lies just medial to the putamen; they are called the lenticular nuclei because they look like curved pieces fitting together like lenses. The globus pallidus has two subdivisions, the external and internal segments, which are lateral and medial, respectively. These nuclei are depicted in a frontal section of the brain in [link].

Frontal Section of Cerebral Cortex and Basal Nuclei

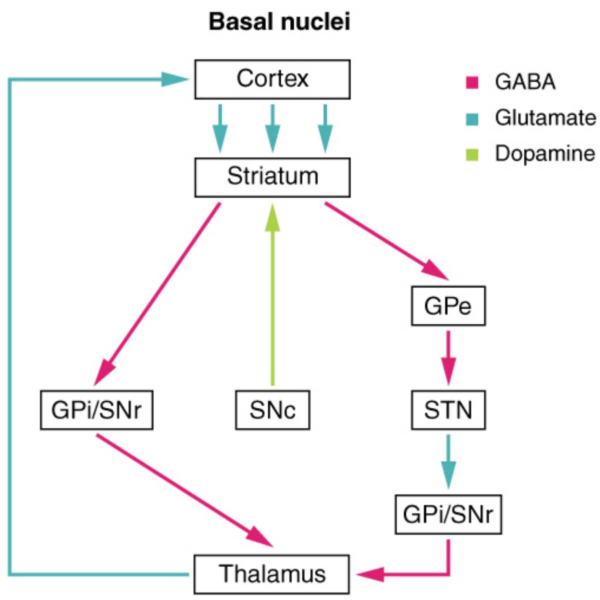


The major components of the basal nuclei, shown in a frontal section of the brain, are the caudate (just lateral to the lateral ventricle), the putamen (inferior to the caudate and separated by the large white-matter structure called the internal

capsule), and the globus pallidus (medial to the putamen).

The basal nuclei in the cerebrum are connected with a few more nuclei in the brain stem that together act as a functional group that forms a motor pathway. Two streams of information processing take place in the basal nuclei. All input to the basal nuclei is from the cortex into the striatum ([link]). The **direct pathway** is the projection of axons from the striatum to the globus pallidus internal segment (GPi) and the **substantia nigra pars reticulata** (SNr). The GPi/SNr then projects to the thalamus, which projects back to the cortex. The **indirect pathway** is the projection of axons from the striatum to the globus pallidus external segment (GPe), then to the subthalamic nucleus (STN), and finally to GPi/SNr. The two streams both target the GPi/SNr, but one has a direct projection and the other goes through a few intervening nuclei. The direct pathway causes the **disinhibition** of the thalamus (inhibition of one cell on a target cell that then inhibits the first cell), whereas the indirect pathway causes, or reinforces, the normal inhibition of the thalamus. The thalamus then can either excite the cortex (as a result of the direct pathway) or fail to excite the cortex (as a result of the indirect pathway).

Connections of Basal Nuclei



Input to the basal nuclei is from the cerebral cortex, which is an

excitatory connection releasing glutamate as a neurotransmitter. This input is to the striatum, or the caudate and putamen. In the direct pathway, the striatum projects to the internal segment of the globus pallidus and the substantia nigra pars reticulata (GPi/SNr). This is an inhibitory pathway, in which GABA is released at the synapse, and the target cells are hyperpolarized and less likely to fire. The output from the basal nuclei is to the thalamus, which is an inhibitory projection using GABA.

The switch between the two pathways is the **substantia nigra pars compacta**, which projects to the striatum and releases the neurotransmitter dopamine. Dopamine receptors are either excitatory (D1-type receptors) or inhibitory (D2-type receptors). The direct pathway is activated by dopamine, and the indirect pathway is inhibited by dopamine. When the substantia nigra pars compacta is firing, it signals to the basal nuclei that the body is in an active state, and movement will be more likely. When the substantia nigra pars compacta is silent, the body is in a passive state, and movement is inhibited. To illustrate this situation, while a student is sitting listening to a lecture, the substantia nigra pars compacta would be silent and the student less likely to get up and walk around. Likewise, while the professor is lecturing, and walking around at the front of the classroom, the professor’s substantia nigra pars compacta would be active, in keeping with his or her activity level.

Note:

Watch this video to learn about the basal nuclei (also known as the basal ganglia), which have two pathways that process information within the cerebrum. As shown in this video, the direct pathway is the shorter pathway through the system that results in increased activity in the cerebral cortex and increased motor activity. The direct pathway is described as resulting in “disinhibition” of the thalamus. What does disinhibition mean? What are the two neurons doing individually to cause this?

Note:

Watch this video to learn about the basal nuclei (also known as the basal ganglia), which have two pathways that process information within the cerebrum. As shown in this video, the indirect pathway is the longer pathway through the system that results in decreased activity in the cerebral cortex, and therefore less motor activity. The indirect pathway has an extra couple of connections in it, including disinhibition of the subthalamic nucleus. What is the end result on the thalamus, and therefore on movement initiated by the cerebral cortex?

Note:
Everyday Connections
The Myth of Left Brain/Right Brain
There is a persistent myth that people are “right-brained” or “left-brained,” which is an oversimplification of an important concept about the cerebral hemispheres. There is some lateralization of function, in which the left side of the brain is devoted to language function and the right side is devoted to spatial and nonverbal reasoning. Whereas these functions are predominantly associated with those sides of the brain, there is no monopoly by either side on these functions. Many pervasive functions, such as language, are distributed globally around the cerebrum.
Some of the support for this misconception has come from studies of split brains. A drastic way to deal with a rare and devastating neurological condition (intractable epilepsy) is to separate the two hemispheres of the brain. After sectioning the corpus callosum, a split-brained patient will have trouble producing verbal responses on the basis of sensory information processed on the right side of the cerebrum, leading to the idea that the left side is responsible for language function.
However, there are well-documented cases of language functions lost from damage to the right side of the brain. The deficits seen in damage to the left side of the brain are classified as aphasia, a loss of speech function; damage on the right side can affect the use of language. Right-side damage can result in a loss of ability to understand figurative aspects of speech, such as jokes, irony, or metaphors. Nonverbal aspects of speech can be affected by damage to the right side, such as facial expression or body language, and right-side damage can lead to a “flat affect” in speech, or a loss of emotional expression in speech—sounding like a robot when talking.

The Diencephalon

The diencephalon is the one region of the adult brain that retains its name from embryologic development. The etymology of the word diencephalon translates to “through brain.” It is the connection between the cerebrum and the rest of the nervous system, with one exception. The rest of the brain, the

spinal cord, and the PNS all send information to the cerebrum through the diencephalon. Output from the cerebrum passes through the diencephalon. The single exception is the system associated with **olfaction**, or the sense of smell, which connects directly with the cerebrum. In the earliest vertebrate species, the cerebrum was not much more than olfactory bulbs that received peripheral information about the chemical environment (to call it smell in these organisms is imprecise because they lived in the ocean).

The diencephalon is deep beneath the cerebrum and constitutes the walls of the third ventricle. The diencephalon can be described as any region of the brain with “thalamus” in its name. The two major regions of the diencephalon are the thalamus itself and the hypothalamus ([link]). There are other structures, such as the **epithalamus**, which contains the pineal gland, or the **subthalamus**, which includes the subthalamic nucleus that is part of the basal nuclei.

Thalamus

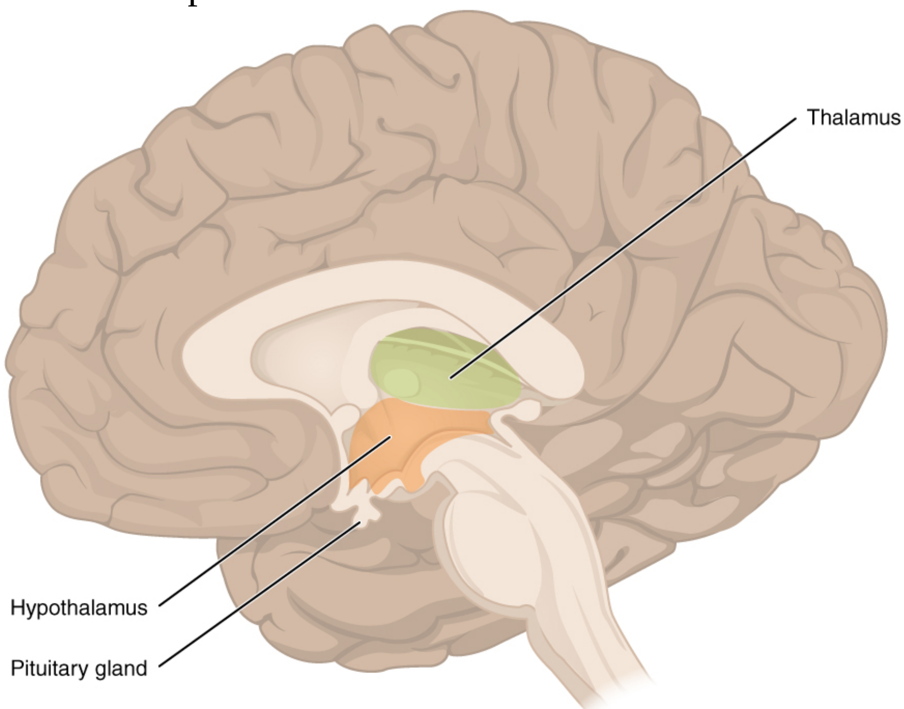
The **thalamus** is a collection of nuclei that relay information between the cerebral cortex and the periphery, spinal cord, or brain stem. All sensory information, except for the sense of smell, passes through the thalamus before processing by the cortex. Axons from the peripheral sensory organs, or intermediate nuclei, synapse in the thalamus, and thalamic neurons project directly to the cerebrum. It is a requisite synapse in any sensory pathway, except for olfaction. The thalamus does not just pass the information on, it also processes that information. For example, the portion of the thalamus that receives visual information will influence what visual stimuli are important, or what receives attention.

The cerebrum also sends information down to the thalamus, which usually communicates motor commands. This involves interactions with the cerebellum and other nuclei in the brain stem. The cerebrum interacts with the basal nuclei, which involves connections with the thalamus. The primary output of the basal nuclei is to the thalamus, which relays that output to the cerebral cortex. The cortex also sends information to the thalamus that will then influence the effects of the basal nuclei.

Hypothalamus

Inferior and slightly anterior to the thalamus is the **hypothalamus**, the other major region of the diencephalon. The hypothalamus is a collection of nuclei that are largely involved in regulating homeostasis. The hypothalamus is the executive region in charge of the autonomic nervous system and the endocrine system through its regulation of the anterior pituitary gland. Other parts of the hypothalamus are involved in memory and emotion as part of the limbic system.

The Diencephalon



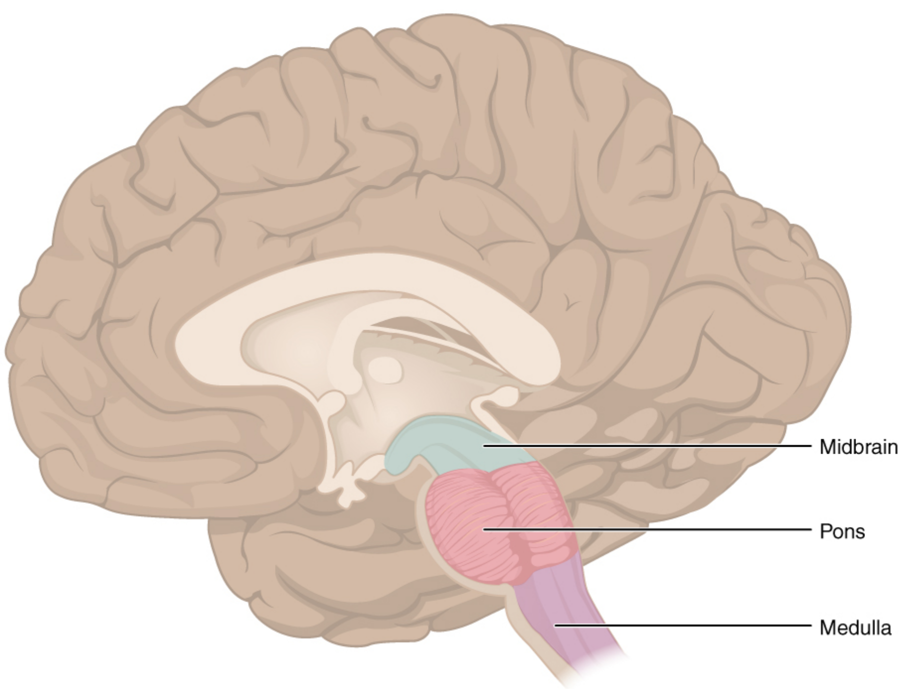
The diencephalon is composed primarily of the thalamus and hypothalamus, which together define the walls of the third ventricle. The thalami are two elongated, ovoid structures on either side of the midline that make contact in the middle. The hypothalamus is inferior and anterior to the thalamus, culminating in a sharp angle to which the pituitary gland is attached.

Brain Stem

The midbrain and hindbrain (composed of the pons and the medulla) are collectively referred to as the brain stem ([link]). The structure emerges from the ventral surface of the forebrain as a tapering cone that connects the brain to the spinal cord. Attached to the brain stem, but considered a separate region of the adult brain, is the cerebellum. The midbrain coordinates sensory representations of the visual, auditory, and somatosensory perceptual spaces. The pons is the main connection with the cerebellum. The pons and the medulla regulate several crucial functions, including the cardiovascular and respiratory systems and rates.

The cranial nerves connect through the brain stem and provide the brain with the sensory input and motor output associated with the head and neck, including most of the special senses. The major ascending and descending pathways between the spinal cord and brain, specifically the cerebrum, pass through the brain stem.

The Brain Stem



The brain stem comprises three regions: the midbrain, the pons, and the medulla.

Midbrain

One of the original regions of the embryonic brain, the midbrain is a small region between the thalamus and pons. It is separated into the **tectum** and **tegmentum**, from the Latin words for roof and floor, respectively. The cerebral aqueduct passes through the center of the midbrain, such that these regions are the roof and floor of that canal.

The tectum is composed of four bumps known as the colliculi (singular = colliculus), which means “little hill” in Latin. The **inferior colliculus** is the inferior pair of these enlargements and is part of the auditory brain stem pathway. Neurons of the inferior colliculus project to the thalamus, which then sends auditory information to the cerebrum for the conscious perception of sound. The **superior colliculus** is the superior pair and combines sensory information about visual space, auditory space, and somatosensory space. Activity in the superior colliculus is related to orienting the eyes to a sound or touch stimulus. If you are walking along the sidewalk on campus and you hear chirping, the superior colliculus coordinates that information with your awareness of the visual location of the tree right above you. That is the correlation of auditory and visual maps. If you suddenly feel something wet fall on your head, your superior colliculus integrates that with the auditory and visual maps and you know that the chirping bird just relieved itself on you. You want to look up to see the culprit, but do not.

The tegmentum is continuous with the gray matter of the rest of the brain stem. Throughout the midbrain, pons, and medulla, the tegmentum contains the nuclei that receive and send information through the cranial nerves, as well as regions that regulate important functions such as those of the cardiovascular and respiratory systems.

Pons

The word pons comes from the Latin word for bridge. It is visible on the anterior surface of the brain stem as the thick bundle of white matter attached to the cerebellum. The pons is the main connection between the cerebellum and the brain stem. The bridge-like white matter is only the anterior surface of the pons; the gray matter beneath that is a continuation of the tegmentum from the midbrain. Gray matter in the tegmentum region of the pons contains neurons receiving descending input from the forebrain that is sent to the cerebellum.

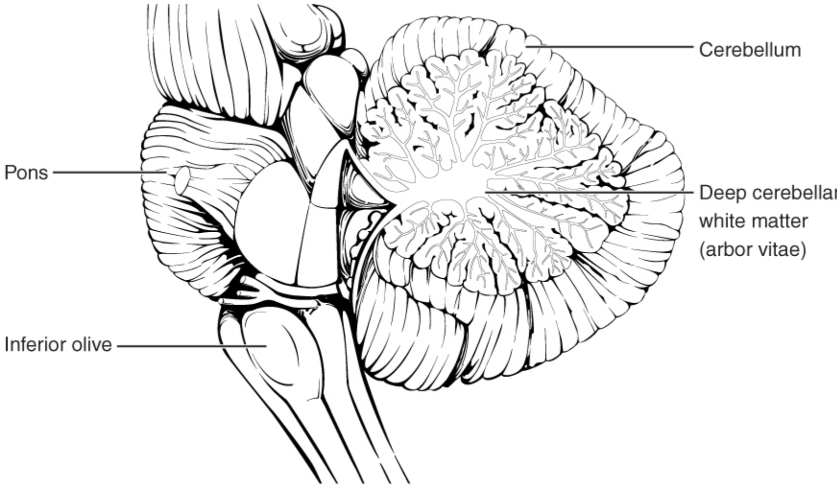
Medulla

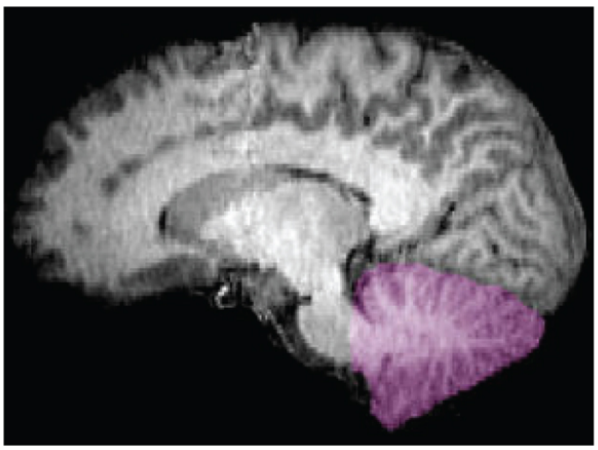
The medulla is the region known as the myelencephalon in the embryonic brain. The initial portion of the name, “myel,” refers to the significant white matter found in this region—especially on its exterior, which is continuous with the white matter of the spinal cord. The tegmentum of the midbrain and pons continues into the medulla because this gray matter is responsible for processing cranial nerve information. A diffuse region of gray matter throughout the brain stem, known as the **reticular formation**, is related to sleep and wakefulness, such as general brain activity and attention.

The Cerebellum

The **cerebellum**, as the name suggests, is the “little brain.” It is covered in gyri and sulci like the cerebrum, and looks like a miniature version of that part of the brain ([link]). The cerebellum is largely responsible for comparing information from the cerebrum with sensory feedback from the periphery through the spinal cord. It accounts for approximately 10 percent of the mass of the brain.

The Cerebellum





The cerebellum is situated on the posterior surface of the brain stem. Descending input from the cerebellum enters through the large white matter structure of the pons. Ascending input from the periphery and spinal cord enters through the fibers of the inferior olive. Output goes to the midbrain, which sends a descending signal to the spinal cord.

Descending fibers from the cerebrum have branches that connect to neurons in the pons. Those neurons project into the cerebellum, providing a copy of

motor commands sent to the spinal cord. Sensory information from the periphery, which enters through spinal or cranial nerves, is copied to a nucleus in the medulla known as the **inferior olive**. Fibers from this nucleus enter the cerebellum and are compared with the descending commands from the cerebrum. If the primary motor cortex of the frontal lobe sends a command down to the spinal cord to initiate walking, a copy of that instruction is sent to the cerebellum. Sensory feedback from the muscles and joints, proprioceptive information about the movements of walking, and sensations of balance are sent to the cerebellum through the inferior olive and the cerebellum compares them. If walking is not coordinated, perhaps because the ground is uneven or a strong wind is blowing, then the cerebellum sends out a corrective command to compensate for the difference between the original cortical command and the sensory feedback. The output of the cerebellum is into the midbrain, which then sends a descending input to the spinal cord to correct the messages going to skeletal muscles.

The Spinal Cord

The description of the CNS is concentrated on the structures of the brain, but the spinal cord is another major organ of the system. Whereas the brain develops out of expansions of the neural tube into primary and then secondary vesicles, the spinal cord maintains the tube structure and is only specialized into certain regions. As the spinal cord continues to develop in the newborn, anatomical features mark its surface. The anterior midline is marked by the **anterior median fissure**, and the posterior midline is marked by the **posterior median sulcus**. Axons enter the posterior side through the **dorsal (posterior) nerve root**, which marks the **posterolateral sulcus** on either side. The axons emerging from the anterior side do so through the **ventral (anterior) nerve root**. Note that it is common to see the terms dorsal (dorsal = “back”) and ventral (ventral = “belly”) used interchangeably with posterior and anterior, particularly in reference to nerves and the structures of the spinal cord. You should learn to be comfortable with both.

On the whole, the posterior regions are responsible for sensory functions and the anterior regions are associated with motor functions. This comes

from the initial development of the spinal cord, which is divided into the **basal plate** and the **alar plate**. The basal plate is closest to the ventral midline of the neural tube, which will become the anterior face of the spinal cord and gives rise to motor neurons. The alar plate is on the dorsal side of the neural tube and gives rise to neurons that will receive sensory input from the periphery.

The length of the spinal cord is divided into regions that correspond to the regions of the vertebral column. The name of a spinal cord region corresponds to the level at which spinal nerves pass through the intervertebral foramina. Immediately adjacent to the brain stem is the cervical region, followed by the thoracic, then the lumbar, and finally the sacral region. The spinal cord is not the full length of the vertebral column because the spinal cord does not grow significantly longer after the first or second year, but the skeleton continues to grow. The nerves that emerge from the spinal cord pass through the intervertebral formina at the respective levels. As the vertebral column grows, these nerves grow with it and result in a long bundle of nerves that resembles a horse's tail and is named the **cauda equina**. The sacral spinal cord is at the level of the upper lumbar vertebral bones. The spinal nerves extend from their various levels to the proper level of the vertebral column.

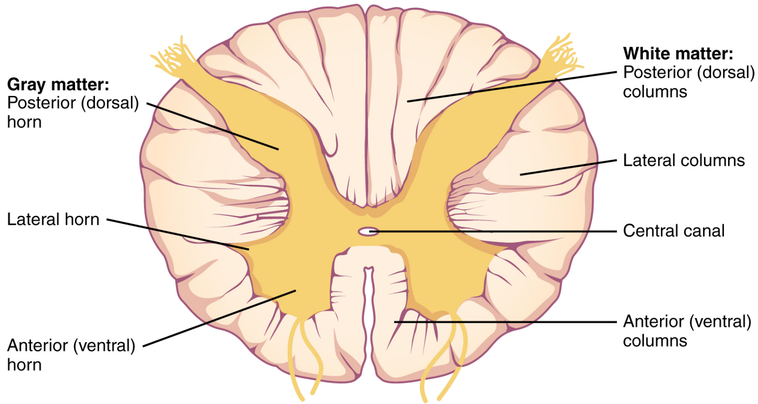
Gray Horns

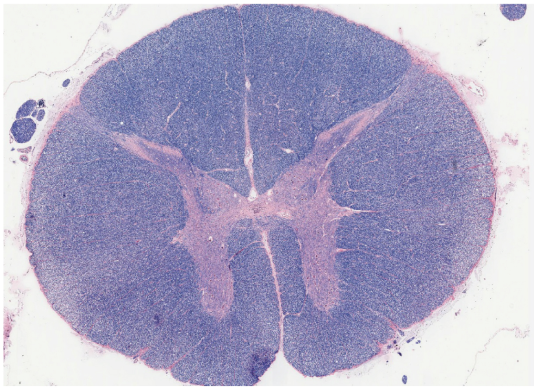
In cross-section, the gray matter of the spinal cord has the appearance of an ink-blot test, with the spread of the gray matter on one side replicated on the other—a shape reminiscent of a bulbous capital "H." As shown in [link], the gray matter is subdivided into regions that are referred to as horns. The **posterior horn** is responsible for sensory processing. The **anterior horn** sends out motor signals to the skeletal muscles. The **lateral horn**, which is only found in the thoracic, upper lumbar, and sacral regions, is the central component of the sympathetic division of the autonomic nervous system.

Some of the largest neurons of the spinal cord are the multipolar motor neurons in the anterior horn. The fibers that cause contraction of skeletal muscles are the axons of these neurons. The motor neuron that causes

contraction of the big toe, for example, is located in the sacral spinal cord. The axon that has to reach all the way to the belly of that muscle may be a meter in length. The neuronal cell body that maintains that long fiber must be quite large, possibly several hundred micrometers in diameter, making it one of the largest cells in the body.

Cross-section of Spinal Cord





The cross-section of a thoracic spinal cord segment shows the posterior, anterior, and lateral horns of gray matter, as well as the posterior, anterior, and lateral columns of white matter. LM × 40. (Micrograph provided by the Regents of University of Michigan Medical School © 2012)

White Columns

Just as the gray matter is separated into horns, the white matter of the spinal cord is separated into columns. **Ascending tracts** of nervous system fibers in these columns carry sensory information up to the brain, whereas **descending tracts** carry motor commands from the brain. Looking at the spinal cord longitudinally, the columns extend along its length as continuous bands of white matter. Between the two posterior horns of gray matter are the **posterior columns**. Between the two anterior horns, and bounded by the axons of motor neurons emerging from that gray matter area, are the **anterior columns**. The white matter on either side of the spinal cord, between the posterior horn and the axons of the anterior horn neurons, are the **lateral columns**. The posterior columns are composed of axons of ascending tracts. The anterior and lateral columns are composed of many different groups of axons of both ascending and descending tracts—the latter carrying motor commands down from the brain to the spinal cord to control output to the periphery.

Note:

Watch this video to learn about the gray matter of the spinal cord that receives input from fibers of the dorsal (posterior) root and sends information out through the fibers of the ventral (anterior) root. As discussed in this video, these connections represent the interactions of the CNS with peripheral structures for both sensory and motor functions. The cervical and lumbar spinal cords have enlargements as a result of larger populations of neurons. What are these enlargements responsible for?

Note:

Disorders of the…

Basal Nuclei

Parkinson’s disease is a disorder of the basal nuclei, specifically of the substantia nigra, that demonstrates the effects of the direct and indirect pathways. Parkinson’s disease is the result of neurons in the substantia nigra pars compacta dying. These neurons release dopamine into the striatum. Without that modulatory influence, the basal nuclei are stuck in the indirect pathway, without the direct pathway being activated. The direct pathway is responsible for increasing cortical movement commands. The increased activity of the indirect pathway results in the hypokinetic disorder of Parkinson’s disease.

Parkinson’s disease is neurodegenerative, meaning that neurons die that cannot be replaced, so there is no cure for the disorder. Treatments for Parkinson’s disease are aimed at increasing dopamine levels in the striatum. Currently, the most common way of doing that is by providing the amino acid L-DOPA, which is a precursor to the neurotransmitter dopamine and can cross the blood-brain barrier. With levels of the precursor elevated, the remaining cells of the substantia nigra pars compacta can make more neurotransmitter and have a greater effect. Unfortunately, the patient will become less responsive to L-DOPA treatment as time progresses, and it can cause increased dopamine levels elsewhere in the brain, which are associated with psychosis or schizophrenia.

Note:

Visit this site for a thorough explanation of Parkinson’s disease.

Note:

Compared with the nearest evolutionary relative, the chimpanzee, the human has a brain that is huge. At a point in the past, a common ancestor gave rise to the two species of humans and chimpanzees. That evolutionary history is long and is still an area of intense study. But something happened to increase the size of the human brain relative to the chimpanzee. Read this article in which the author explores the current understanding of why this happened.

According to one hypothesis about the expansion of brain size, what tissue might have been sacrificed so energy was available to grow our larger brain? Based on what you know about that tissue and nervous tissue, why would there be a trade-off between them in terms of energy use?

Chapter Review

The adult brain is separated into four major regions: the cerebrum, the diencephalon, the brain stem, and the cerebellum. The cerebrum is the largest portion and contains the cerebral cortex and subcortical nuclei. It is divided into two halves by the longitudinal fissure.

The cortex is separated into the frontal, parietal, temporal, and occipital lobes. The frontal lobe is responsible for motor functions, from planning movements through executing commands to be sent to the spinal cord and periphery. The most anterior portion of the frontal lobe is the prefrontal cortex, which is associated with aspects of personality through its influence on motor responses in decision-making.

The other lobes are responsible for sensory functions. The parietal lobe is where somatosensation is processed. The occipital lobe is where visual processing begins, although the other parts of the brain can contribute to visual function. The temporal lobe contains the cortical area for auditory processing, but also has regions crucial for memory formation.

Nuclei beneath the cerebral cortex, known as the subcortical nuclei, are responsible for augmenting cortical functions. The basal nuclei receive input from cortical areas and compare it with the general state of the individual through the activity of a dopamine-releasing nucleus. The output influences the activity of part of the thalamus that can then increase or decrease cortical activity that often results in changes to motor commands. The basal forebrain is responsible for modulating cortical activity in attention and memory. The limbic system includes deep cerebral nuclei that are responsible for emotion and memory.

The diencephalon includes the thalamus and the hypothalamus, along with some other structures. The thalamus is a relay between the cerebrum and the rest of the nervous system. The hypothalamus coordinates homeostatic functions through the autonomic and endocrine systems.

The brain stem is composed of the midbrain, pons, and medulla. It controls the head and neck region of the body through the cranial nerves. There are control centers in the brain stem that regulate the cardiovascular and respiratory systems.

The cerebellum is connected to the brain stem, primarily at the pons, where it receives a copy of the descending input from the cerebrum to the spinal cord. It can compare this with sensory feedback input through the medulla and send output through the midbrain that can correct motor commands for coordination.

Interactive Link Questions

Exercise:

Problem:

Watch this video to learn about the basal nuclei (also known as the basal ganglia), which have two pathways that process information within the cerebrum. As shown in this video, the direct pathway is the shorter pathway through the system that results in increased activity in the cerebral cortex and increased motor activity. The direct pathway is described as resulting in “disinhibition” of the thalamus. What does disinhibition mean? What are the two neurons doing individually to cause this?

Solution:

Both cells are inhibitory. The first cell inhibits the second one. Therefore, the second cell can no longer inhibit its target. This is disinhibition of that target across two synapses.

Exercise:

Problem:

Watch this video to learn about the basal nuclei (also known as the basal ganglia), which have two pathways that process information within the cerebrum. As shown in this video, the indirect pathway is the longer pathway through the system that results in decreased activity in the cerebral cortex, and therefore less motor activity. The indirect pathway has an extra couple of connections in it, including disinhibition of the subthalamic nucleus. What is the end result on the thalamus, and therefore on movement initiated by the cerebral cortex?

Solution:

By disinhibiting the subthalamic nucleus, the indirect pathway increases excitation of the globus pallidus internal segment. That, in turn, inhibits the thalamus, which is the opposite effect of the direct pathway that disinhibits the thalamus.

Exercise:

Problem:

Watch this video to learn about the gray matter of the spinal cord that receives input from fibers of the dorsal (posterior) root and sends information out through the fibers of the ventral (anterior) root. As discussed in this video, these connections represent the interactions of the CNS with peripheral structures for both sensory and motor functions. The cervical and lumbar spinal cords have enlargements as a result of larger populations of neurons. What are these enlargements responsible for?

Solution:

There are more motor neurons in the anterior horns that are responsible for movement in the limbs. The cervical enlargement is for the arms, and the lumbar enlargement is for the legs.

Exercise:

Problem:

Compared with the nearest evolutionary relative, the chimpanzee, the human has a brain that is huge. At a point in the past, a common ancestor gave rise to the two species of humans and chimpanzees. That evolutionary history is long and is still an area of intense study. But something happened to increase the size of the human brain relative to the chimpanzee. Read this article in which the author explores the current understanding of why this happened.

According to one hypothesis about the expansion of brain size, what tissue might have been sacrificed so energy was available to grow our larger brain? Based on what you know about that tissue and nervous tissue, why would there be a trade-off between them in terms of energy use?

Solution:

Energy is needed for the brain to develop and perform higher cognitive functions. That energy is not available for the muscle tissues to develop and function. The hypothesis suggests that humans have larger brains and less muscle mass, and chimpanzees have the smaller brains but more muscle mass.

Review Questions

Exercise:

Problem:

Which lobe of the cerebral cortex is responsible for generating motor commands?

a. temporal
b. parietal
c. occipital
d. frontal

Solution:

D

Exercise:

Problem: What region of the diencephalon coordinates homeostasis?

a. thalamus
b. epithalamus
c. hypothalamus
d. subthalamus

Solution:

C

Exercise:

Problem:

What level of the brain stem is the major input to the cerebellum?

a. midbrain
b. pons
c. medulla
d. spinal cord

Solution:

B

Exercise:

Problem:

What region of the spinal cord contains motor neurons that direct the movement of skeletal muscles?

a. anterior horn
b. posterior horn
c. lateral horn
d. alar plate

Solution:

A

Exercise:

Problem:

Brodmann's areas map different regions of the ________ to particular functions.

a. cerebellum

b. cerebral cortex
c. basal forebrain
d. corpus callosum

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

Damage to specific regions of the cerebral cortex, such as through a stroke, can result in specific losses of function. What functions would likely be lost by a stroke in the temporal lobe?

Solution:

The temporal lobe has sensory functions associated with hearing and vision, as well as being important for memory. A stroke in the temporal lobe can result in specific sensory deficits in these systems (known as agnosias) or losses in memory.

Exercise:

Problem:

Why do the anatomical inputs to the cerebellum suggest that it can compare motor commands and sensory feedback?

Solution:

A copy of descending input from the cerebrum to the spinal cord, through the pons, and sensory feedback from the spinal cord and special senses like balance, through the medulla, both go to the

cerebellum. It can therefore send output through the midbrain that will correct spinal cord control of skeletal muscle movements.

Glossary

alar plate
: developmental region of the spinal cord that gives rise to the posterior horn of the gray matter

amygdala
: nucleus deep in the temporal lobe of the cerebrum that is related to memory and emotional behavior

anterior column
: white matter between the anterior horns of the spinal cord composed of many different groups of axons of both ascending and descending tracts

anterior horn
: gray matter of the spinal cord containing multipolar motor neurons, sometimes referred to as the ventral horn

anterior median fissure
: deep midline feature of the anterior spinal cord, marking the separation between the right and left sides of the cord

ascending tract
: central nervous system fibers carrying sensory information from the spinal cord or periphery to the brain

basal forebrain
: nuclei of the cerebrum related to modulation of sensory stimuli and attention through broad projections to the cerebral cortex, loss of which is related to Alzheimer’s disease

basal nuclei

nuclei of the cerebrum (with a few components in the upper brain stem and diencephalon) that are responsible for assessing cortical movement commands and comparing them with the general state of the individual through broad modulatory activity of dopamine neurons; largely related to motor functions, as evidenced through the symptoms of Parkinson’s and Huntington’s diseases

basal plate
developmental region of the spinal cord that gives rise to the lateral and anterior horns of gray matter

Broca’s area
region of the frontal lobe associated with the motor commands necessary for speech production and located only in the cerebral hemisphere responsible for language production, which is the left side in approximately 95 percent of the population

Brodmann’s areas
mapping of regions of the cerebral cortex based on microscopic anatomy that relates specific areas to functional differences, as described by Brodmann in the early 1900s

cauda equina
bundle of spinal nerve roots that descend from the lower spinal cord below the first lumbar vertebra and lie within the vertebral cavity; has the appearance of a horse's tail

caudate
nucleus deep in the cerebrum that is part of the basal nuclei; along with the putamen, it is part of the striatum

central sulcus
surface landmark of the cerebral cortex that marks the boundary between the frontal and parietal lobes

cerebral cortex
outer gray matter covering the forebrain, marked by wrinkles and folds known as gyri and sulci

cerebrum
: region of the adult brain that develops from the telencephalon and is responsible for higher neurological functions such as memory, emotion, and consciousness

cerebellum
: region of the adult brain connected primarily to the pons that developed from the metencephalon (along with the pons) and is largely responsible for comparing information from the cerebrum with sensory feedback from the periphery through the spinal cord

cerebral hemisphere
: one half of the bilaterally symmetrical cerebrum

corpus callosum
: large white matter structure that connects the right and left cerebral hemispheres

descending tract
: central nervous system fibers carrying motor commands from the brain to the spinal cord or periphery

direct pathway
: connections within the basal nuclei from the striatum to the globus pallidus internal segment and substantia nigra pars reticulata that disinhibit the thalamus to increase cortical control of movement

disinhibition
: disynaptic connection in which the first synapse inhibits the second cell, which then stops inhibiting the final target

dorsal (posterior) nerve root
: axons entering the posterior horn of the spinal cord

epithalamus
: region of the diecephalon containing the pineal gland

frontal eye field

region of the frontal lobe associated with motor commands to orient the eyes toward an object of visual attention

frontal lobe
region of the cerebral cortex directly beneath the frontal bone of the cranium

globus pallidus
nuclei deep in the cerebrum that are part of the basal nuclei and can be divided into the internal and external segments

gyrus
ridge formed by convolutions on the surface of the cerebrum or cerebellum

hippocampus
gray matter deep in the temporal lobe that is very important for long-term memory formation

hypothalamus
major region of the diencephalon that is responsible for coordinating autonomic and endocrine control of homeostasis

indirect pathway
connections within the basal nuclei from the striatum through the globus pallidus external segment and subthalamic nucleus to the globus pallidus internal segment/substantia nigra pars compacta that result in inhibition of the thalamus to decrease cortical control of movement

inferior colliculus
half of the midbrain tectum that is part of the brain stem auditory pathway

inferior olive
nucleus in the medulla that is involved in processing information related to motor control

kinesthesia
general sensory perception of movement of the body

lateral column
white matter of the spinal cord between the posterior horn on one side and the axons from the anterior horn on the same side; composed of many different groups of axons, of both ascending and descending tracts, carrying motor commands to and from the brain

lateral horn
region of the spinal cord gray matter in the thoracic, upper lumbar, and sacral regions that is the central component of the sympathetic division of the autonomic nervous system

lateral sulcus
surface landmark of the cerebral cortex that marks the boundary between the temporal lobe and the frontal and parietal lobes

limbic cortex
collection of structures of the cerebral cortex that are involved in emotion, memory, and behavior and are part of the larger limbic system

limbic system
structures at the edge (limit) of the boundary between the forebrain and hindbrain that are most associated with emotional behavior and memory formation

longitudinal fissure
large separation along the midline between the two cerebral hemispheres

occipital lobe
region of the cerebral cortex directly beneath the occipital bone of the cranium

olfaction

special sense responsible for smell, which has a unique, direct connection to the cerebrum

parietal lobe
: region of the cerebral cortex directly beneath the parietal bone of the cranium

parieto-occipital sulcus
: groove in the cerebral cortex representing the border between the parietal and occipital cortices

postcentral gyrus
: primary motor cortex located in the frontal lobe of the cerebral cortex

posterior columns
: white matter of the spinal cord that lies between the posterior horns of the gray matter, sometimes referred to as the dorsal column; composed of axons of ascending tracts that carry sensory information up to the brain

posterior horn
: gray matter region of the spinal cord in which sensory input arrives, sometimes referred to as the dorsal horn

posterior median sulcus
: midline feature of the posterior spinal cord, marking the separation between right and left sides of the cord

posterolateral sulcus
: feature of the posterior spinal cord marking the entry of posterior nerve roots and the separation between the posterior and lateral columns of the white matter

precentral gyrus
: ridge just posterior to the central sulcus, in the parietal lobe, where somatosensory processing initially takes place in the cerebrum

prefrontal lobe

specific region of the frontal lobe anterior to the more specific motor function areas, which can be related to the early planning of movements and intentions to the point of being personality-type functions

premotor area
: region of the frontal lobe responsible for planning movements that will be executed through the primary motor cortex

proprioception
: general sensory perceptions providing information about location and movement of body parts; the “sense of the self”

putamen
: nucleus deep in the cerebrum that is part of the basal nuclei; along with the caudate, it is part of the striatum

reticular formation
: diffuse region of gray matter throughout the brain stem that regulates sleep, wakefulness, and states of consciousness

somatosensation
: general senses related to the body, usually thought of as the senses of touch, which would include pain, temperature, and proprioception

striatum
: the caudate and putamen collectively, as part of the basal nuclei, which receive input from the cerebral cortex

subcortical nucleus
: all the nuclei beneath the cerebral cortex, including the basal nuclei and the basal forebrain

substantia nigra pars compacta
: nuclei within the basal nuclei that release dopamine to modulate the function of the striatum; part of the motor pathway

substantia nigra pars reticulata

nuclei within the basal nuclei that serve as an output center of the nuclei; part of the motor pathway

subthalamus
: nucleus within the basal nuclei that is part of the indirect pathway

sulcus
: groove formed by convolutions in the surface of the cerebral cortex

superior colliculus
: half of the midbrain tectum that is responsible for aligning visual, auditory, and somatosensory spatial perceptions

tectum
: region of the midbrain, thought of as the roof of the cerebral aqueduct, which is subdivided into the inferior and superior colliculi

tegmentum
: region of the midbrain, thought of as the floor of the cerebral aqueduct, which continues into the pons and medulla as the floor of the fourth ventricle

temporal lobe
: region of the cerebral cortex directly beneath the temporal bone of the cranium

thalamus
: major region of the diencephalon that is responsible for relaying information between the cerebrum and the hindbrain, spinal cord, and periphery

ventral (anterior) nerve root
: axons emerging from the anterior or lateral horns of the spinal cord

The Peripheral Nervous System
By the end of this section, you will be able to:

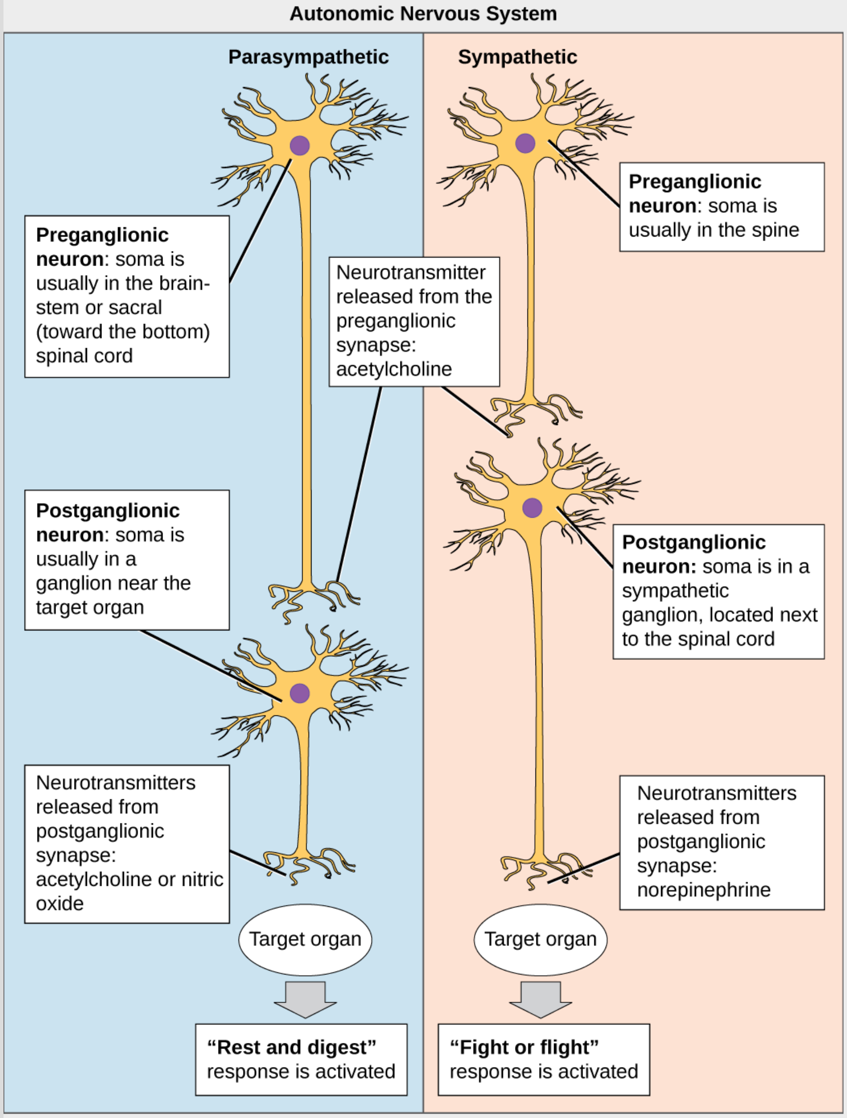
- Describe the organization and functions of the sympathetic and parasympathetic nervous systems
- Describe the organization and function of the sensory-somatic nervous system

The peripheral nervous system (PNS) is the connection between the central nervous system and the rest of the body. The CNS is like the power plant of the nervous system. It creates the signals that control the functions of the body. The PNS is like the wires that go to individual houses. Without those "wires," the signals produced by the CNS could not control the body (and the CNS would not be able to receive sensory information from the body either).

The PNS can be broken down into the **autonomic nervous system**, which controls bodily functions without conscious control, and the **sensory-somatic nervous system**, which transmits sensory information from the skin, muscles, and sensory organs to the CNS and sends motor commands from the CNS to the muscles.

Autonomic Nervous System

Note:
Art Connection



In the autonomic nervous system, a preganglionic neuron of the CNS synapses with a postganglionic neuron of the PNS. The postganglionic neuron, in turn, acts on a target organ. Autonomic responses are mediated by the sympathetic and the parasympathetic systems, which are antagonistic to one another. The sympathetic system activates the "fight or flight" response, while the parasympathetic system activates the "rest and digest" response.

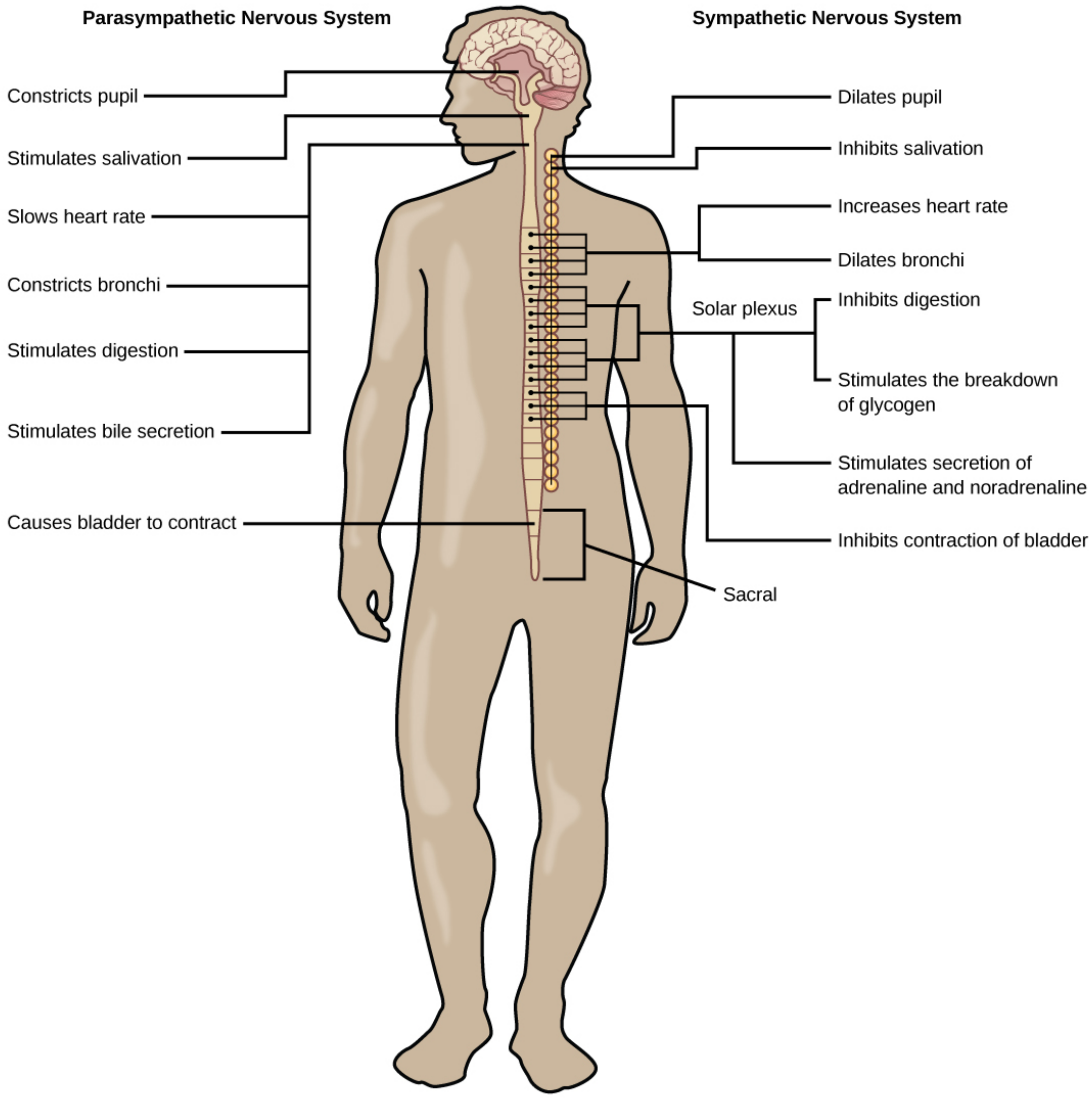
Which of the following statements is false?

a. The parasympathetic pathway is responsible for resting the body, while the sympathetic pathway is responsible for preparing for an emergency.
b. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
c. Slowing of the heartbeat is a parasympathetic response.
d. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.

The autonomic nervous system serves as the relay between the CNS and the internal organs. It controls the lungs, the heart, smooth muscle, and exocrine and endocrine glands. The autonomic nervous system controls these organs largely without conscious control; it can continuously monitor the conditions of these different systems and implement changes as needed. Signaling to the target tissue usually involves two synapses: a preganglionic neuron (originating in the CNS) synapses to a neuron in a ganglion that, in turn, synapses on the target organ, as illustrated in [link]. There are two divisions of the autonomic nervous system that often have opposing effects: the sympathetic nervous system and the parasympathetic nervous system.

Sympathetic Nervous System

The **sympathetic nervous system** is responsible for the “fight or flight” response that occurs when an animal encounters a dangerous situation. One way to remember this is to think of the surprise a person feels when encountering a snake (“snake” and “sympathetic” both begin with “s”). Examples of functions controlled by the sympathetic nervous system include an accelerated heart rate and inhibited digestion. These functions help prepare an organism’s body for the physical strain required to escape a potentially dangerous situation or to fend off a predator.



The sympathetic and parasympathetic nervous systems often have opposing effects on target organs.

Most preganglionic neurons in the sympathetic nervous system originate in the spinal cord, as illustrated in [link]. The axons of these neurons release **acetylcholine** on postganglionic neurons within sympathetic ganglia (the sympathetic ganglia form a chain that extends alongside the spinal cord). The acetylcholine activates the postganglionic neurons. Postganglionic neurons then release **norepinephrine** onto target organs. As anyone who has ever felt a rush before a big test, speech, or athletic event can attest, the

effects of the sympathetic nervous system are quite pervasive. This is both because one preganglionic neuron synapses on multiple postganglionic neurons, amplifying the effect of the original synapse, and because the adrenal gland also releases norepinephrine (and the closely related hormone epinephrine) into the blood stream. The physiological effects of this norepinephrine release include dilating the trachea and bronchi (making it easier for the animal to breathe), increasing heart rate, and moving blood from the skin to the heart, muscles, and brain (so the animal can think and run). The strength and speed of the sympathetic response helps an organism avoid danger, and scientists have found evidence that it may also increase LTP—allowing the animal to remember the dangerous situation and avoid it in the future.

Parasympathetic Nervous System

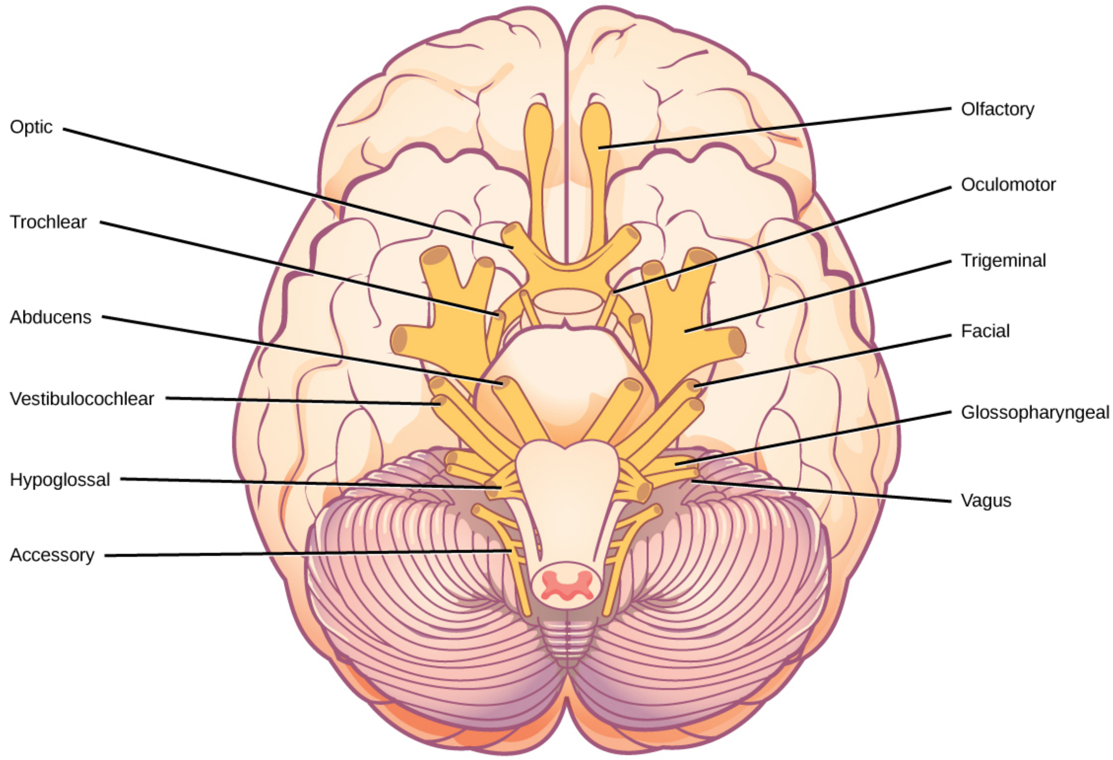
While the sympathetic nervous system is activated in stressful situations, the **parasympathetic nervous system** allows an animal to “rest and digest.” One way to remember this is to think that during a restful situation like a picnic, the parasympathetic nervous system is in control (“picnic” and “parasympathetic” both start with “p”). Parasympathetic preganglionic neurons have cell bodies located in the brainstem and in the sacral (toward the bottom) spinal cord, as shown in [link]. The axons of the preganglionic neurons release acetylcholine on the postganglionic neurons, which are generally located very near the target organs. Most postganglionic neurons release acetylcholine onto target organs, although some release nitric oxide.

The parasympathetic nervous system resets organ function after the sympathetic nervous system is activated (the common adrenaline dump you feel after a ‘fight-or-flight’ event). Effects of acetylcholine release on target organs include slowing of heart rate, lowered blood pressure, and stimulation of digestion.

Sensory-Somatic Nervous System

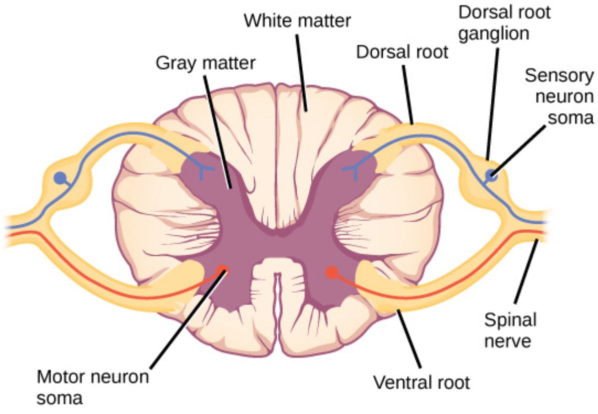
The sensory-somatic nervous system is made up of cranial and spinal nerves and contains both sensory and motor neurons. Sensory neurons transmit sensory information from the skin, skeletal muscle, and sensory organs to the CNS. Motor neurons transmit messages about desired movement from the CNS to the muscles to make them contract. Without its sensory-somatic nervous system, an animal would be unable to process any information about its environment (what it sees, feels, hears, and so on) and could not control motor movements. Unlike the autonomic nervous system, which has two synapses between the CNS and the target organ, sensory and motor neurons have only one synapse—one ending of the neuron is at the organ and the other directly contacts a CNS neuron. Acetylcholine is the main neurotransmitter released at these synapses.

Humans have 12 **cranial nerves**, nerves that emerge from or enter the skull (cranium), as opposed to the spinal nerves, which emerge from the vertebral column. Each cranial nerve is accorded a name, which are detailed in [link]. Some cranial nerves transmit only sensory information. For example, the olfactory nerve transmits information about smells from the nose to the brainstem. Other cranial nerves transmit almost solely motor information. For example, the oculomotor nerve controls the opening and closing of the eyelid and some eye movements. Other cranial nerves contain a mix of sensory and motor fibers. For example, the glossopharyngeal nerve has a role in both taste (sensory) and swallowing (motor).



The human brain contains 12 cranial nerves that receive sensory input and control motor output for the head and neck.

Spinal nerves transmit sensory and motor information between the spinal cord and the rest of the body. Each of the 31 spinal nerves (in humans) contains both sensory and motor axons. The sensory neuron cell bodies are grouped in structures called dorsal root ganglia and are shown in [link]. Each sensory neuron has one projection—with a sensory receptor ending in skin, muscle, or sensory organs—and another that synapses with a neuron in the dorsal spinal cord. Motor neurons have cell bodies in the ventral gray matter of the spinal cord that project to muscle through the ventral root. These neurons are usually stimulated by interneurons within the spinal cord but are sometimes directly stimulated by sensory neurons.



Spinal nerves contain both sensory and motor axons. The somas of sensory neurons are located in dorsal root ganglia. The somas of motor neurons are found in the ventral portion of the gray matter of the spinal cord.

Section Summary

The peripheral nervous system contains both the autonomic and sensory-somatic nervous systems. The autonomic nervous system provides unconscious control over visceral functions and has two divisions: the sympathetic and parasympathetic nervous systems. The sympathetic nervous system is activated in stressful situations to prepare the animal for a “fight or flight” response. The parasympathetic nervous system is active during restful periods. The sensory-somatic nervous system is made of cranial and spinal nerves that transmit sensory information from skin and muscle to the CNS and motor commands from the CNS to the muscles.

Art Connections

Exercise:

Problem: [link] Which of the following statements is false?

a. The parasympathetic pathway is responsible for relaxing the body, while the sympathetic pathway is responsible for preparing for an emergency.
b. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
c. Slowing of the heartbeat is a parasympathetic response.
d. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.

Solution:

[link] D

Review Questions

Exercise:

Problem: Activation of the sympathetic nervous system causes:

a. increased blood flow into the skin
b. a decreased heart rate
c. an increased heart rate
d. increased digestion

Solution:

C

Exercise:

Problem:

Where are parasympathetic preganglionic cell bodies located?

a. cerebellum
b. brainstem
c. dorsal root ganglia
d. skin

Solution:

B

Exercise:

Problem:_________ is released by motor nerve endings onto muscle.

a. Acetylcholine
b. Norepinephrine
c. Dopamine
d. Serotonin

Solution:

A

Free Response

Exercise:

Problem:

What are the main differences between the sympathetic and parasympathetic branches of the autonomic nervous system?

Solution:

The sympathetic nervous system prepares the body for “fight or flight,” whereas the parasympathetic nervous system allows the body to “rest and digest.” Sympathetic neurons release norepinephrine onto target organs; parasympathetic neurons release acetylcholine. Sympathetic neuron cell bodies are located in sympathetic ganglia. Parasympathetic neuron cell bodies are located in the brainstem and sacral spinal cord. Activation of the sympathetic nervous system increases heart rate and blood pressure and decreases digestion and blood flow to the skin. Activation of the parasympathetic nervous system decreases heart rate and blood pressure and increases digestion and blood flow to the skin.

Exercise:

Problem:

What are the main functions of the sensory-somatic nervous system?

Solution:

The sensory-somatic nervous system transmits sensory information from the skin, muscles, and sensory organs to the CNS. It also sends motor commands from the CNS to the muscles, causing them to contract.

Glossary

acetylcholine
: neurotransmitter released by neurons in the central nervous system and peripheral nervous system

autonomic nervous system
: part of the peripheral nervous system that controls bodily functions

cranial nerve
: sensory and/or motor nerve that emanates from the brain

norepinephrine

neurotransmitter and hormone released by activation of the sympathetic nervous system

parasympathetic nervous system
division of autonomic nervous system that regulates visceral functions during rest and digestion

sensory-somatic nervous system
system of sensory and motor nerves

spinal nerve
nerve projecting between skin or muscle and spinal cord

sympathetic nervous system
division of autonomic nervous system activated during stressful “fight or flight” situations

Nervous System Disorders
By the end of this section, you will be able to:

- Describe the symptoms, potential causes, and treatment of several examples of nervous system disorders

A nervous system that functions correctly is a fantastically complex, well-oiled machine—synapses fire appropriately, muscles move when needed, memories are formed and stored, and emotions are well regulated. Unfortunately, each year millions of people in the United States deal with some sort of nervous system disorder. While scientists have discovered potential causes of many of these diseases, and viable treatments for some, ongoing research seeks to find ways to better prevent and treat all of these disorders.

Neurodegenerative Disorders

Neurodegenerative disorders are illnesses characterized by a loss of nervous system functioning that are usually caused by neuronal death. These diseases generally worsen over time as more and more neurons die. The symptoms of a particular neurodegenerative disease are related to where in the nervous system the death of neurons occurs. Spinocerebellar ataxia, for example, leads to neuronal death in the cerebellum. The death of these neurons causes problems in balance and walking. Neurodegenerative disorders include Huntington's disease, amyotrophic lateral sclerosis, Alzheimer's disease and other types of dementia disorders, and Parkinson's disease. Here, Alzheimer's and Parkinson's disease will be discussed in more depth.

Alzheimer's Disease

Alzheimer's disease is the most common cause of dementia in the elderly. In 2012, an estimated 5.4 million Americans suffered from Alzheimer's disease, and payments for their care are estimated at $200 billion. Roughly one in every eight people age 65 or older has the disease. Due to the aging

of the baby-boomer generation, there are projected to be as many as 13 million Alzheimer’s patients in the United States in the year 2050.

Symptoms of Alzheimer’s disease include disruptive memory loss, confusion about time or place, difficulty planning or executing tasks, poor judgment, and personality changes. Problems smelling certain scents can also be indicative of Alzheimer’s disease and may serve as an early warning sign. Many of these symptoms are also common in people who are aging normally, so it is the severity and longevity of the symptoms that determine whether a person is suffering from Alzheimer’s.

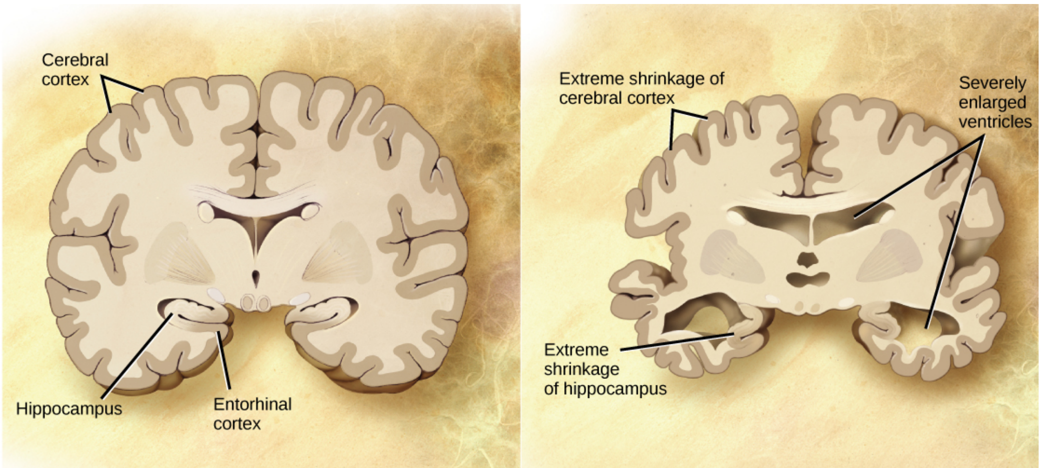
Alzheimer’s disease was named for Alois Alzheimer, a German psychiatrist who published a report in 1911 about a woman who showed severe dementia symptoms. Along with his colleagues, he examined the woman’s brain following her death and reported the presence of abnormal clumps, which are now called amyloid plaques, along with tangled brain fibers called neurofibrillary tangles. Amyloid plaques, neurofibrillary tangles, and an overall shrinking of brain volume are commonly seen in the brains of Alzheimer’s patients. Loss of neurons in the hippocampus is especially severe in advanced Alzheimer’s patients. [link] compares a normal brain to the brain of an Alzheimer’s patient. Many research groups are examining the causes of these hallmarks of the disease.

One form of the disease is usually caused by mutations in one of three known genes. This rare form of early onset Alzheimer’s disease affects fewer than five percent of patients with the disease and causes dementia beginning between the ages of 30 and 60. The more prevalent, late-onset form of the disease likely also has a genetic component. One particular gene, apolipoprotein E (APOE) has a variant (E4) that increases a carrier’s likelihood of getting the disease. Many other genes have been identified that might be involved in the pathology.

Note:
Link to Learning

Visit this website for video links discussing genetics and Alzheimer’s disease.

Unfortunately, there is no cure for Alzheimer’s disease. Current treatments focus on managing the symptoms of the disease. Because decrease in the activity of cholinergic neurons (neurons that use the neurotransmitter acetylcholine) is common in Alzheimer’s disease, several drugs used to treat the disease work by increasing acetylcholine neurotransmission, often by inhibiting the enzyme that breaks down acetylcholine in the synaptic cleft. Other clinical interventions focus on behavioral therapies like psychotherapy, sensory therapy, and cognitive exercises. Since Alzheimer’s disease appears to hijack the normal aging process, research into prevention is prevalent. Smoking, obesity, and cardiovascular problems may be risk factors for the disease, so treatments for those may also help to prevent Alzheimer’s disease. Some studies have shown that people who remain intellectually active by playing games, reading, playing musical instruments, and being socially active in later life have a reduced risk of developing the disease.



Compared to a normal brain (left), the brain from a patient with Alzheimer's disease (right) shows a dramatic neurodegeneration, particularly within the ventricles and hippocampus. (credit: modification of work by "Garrando"/Wikimedia Commons based on original images by ADEAR: "Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging")

Parkinson's Disease

Like Alzheimer's disease, **Parkinson's disease** is a neurodegenerative disease. It was first characterized by James Parkinson in 1817. Each year, 50,000-60,000 people in the United States are diagnosed with the disease. Parkinson's disease causes the loss of dopamine neurons in the substantia nigra, a midbrain structure that regulates movement. Loss of these neurons causes many symptoms including tremor (shaking of fingers or a limb), slowed movement, speech changes, balance and posture problems, and rigid muscles. The combination of these symptoms often causes a characteristic slow hunched shuffling walk, illustrated in [link]. Patients with Parkinson's disease can also exhibit psychological symptoms, such as dementia or emotional problems.

Although some patients have a form of the disease known to be caused by a single mutation, for most patients the exact causes of Parkinson's disease remain unknown: the disease likely results from a combination of genetic and environmental factors (similar to Alzheimer's disease). Post-mortem analysis of brains from Parkinson's patients shows the presence of Lewy bodies—abnormal protein clumps—in dopaminergic neurons. The prevalence of these Lewy bodies often correlates with the severity of the disease.

There is no cure for Parkinson's disease, and treatment is focused on easing symptoms. One of the most commonly prescribed drugs for Parkinson's is L-DOPA, which is a chemical that is converted into dopamine by neurons in the brain. This conversion increases the overall level of dopamine neurotransmission and can help compensate for the loss of dopaminergic neurons in the substantia nigra. Other drugs work by inhibiting the enzyme that breaks down dopamine.

Parkinson's patients often have a characteristic hunched walk.

Neurodevelopmental Disorders

Neurodevelopmental disorders occur when the development of the nervous system is disturbed. There are several different classes of neurodevelopmental disorders. Some, like Down Syndrome, cause intellectual deficits. Others specifically affect communication, learning, or the motor system. Some disorders like autism spectrum disorder and attention deficit/hyperactivity disorder have complex symptoms.

Autism

Autism spectrum disorder (ASD) is a neurodevelopmental disorder. Its severity differs from person to person. Estimates for the prevalence of the disorder have changed rapidly in the past few decades. Current estimates suggest that one in 88 children will develop the disorder. ASD is four times more prevalent in males than females.

Note:
Link to Learning

This video discusses possible reasons why there has been a recent increase in the number of people diagnosed with autism.

A characteristic symptom of ASD is impaired social skills. Children with autism may have difficulty making and maintaining eye contact and reading social cues. They also may have problems feeling empathy for others. Other

symptoms of ASD include repetitive motor behaviors (such as rocking back and forth), preoccupation with specific subjects, strict adherence to certain rituals, and unusual language use. Up to 30 percent of patients with ASD develop epilepsy, and patients with some forms of the disorder (like Fragile X) also have intellectual disability. Because it is a spectrum disorder, other ASD patients are very functional and have good-to-excellent language skills. Many of these patients do not feel that they suffer from a disorder and instead think that their brains just process information differently.

Except for some well-characterized, clearly genetic forms of autism (like Fragile X and Rett's Syndrome), the causes of ASD are largely unknown. Variants of several genes correlate with the presence of ASD, but for any given patient, many different mutations in different genes may be required for the disease to develop. At a general level, ASD is thought to be a disease of "incorrect" wiring. Accordingly, brains of some ASD patients lack the same level of synaptic pruning that occurs in non-affected people. In the 1990s, a research paper linked autism to a common vaccine given to children. This paper was retracted when it was discovered that the author falsified data, and follow-up studies showed no connection between vaccines and autism.

Treatment for autism usually combines behavioral therapies and interventions, along with medications to treat other disorders common to people with autism (depression, anxiety, obsessive compulsive disorder). Although early interventions can help mitigate the effects of the disease, there is currently no cure for ASD.

Attention Deficit Hyperactivity Disorder (ADHD)

Approximately three to five percent of children and adults are affected by **attention deficit/hyperactivity disorder (ADHD)**. Like ASD, ADHD is more prevalent in males than females. Symptoms of the disorder include inattention (lack of focus), executive functioning difficulties, impulsivity, and hyperactivity beyond what is characteristic of the normal developmental stage. Some patients do not have the hyperactive component of symptoms and are diagnosed with a subtype of ADHD: attention deficit

disorder (ADD). Many people with ADHD also show comorbitity, in that they develop secondary disorders in addition to ADHD. Examples include depression or obsessive compulsive disorder (OCD). [link] provides some statistics concerning comorbidity with ADHD.

The cause of ADHD is unknown, although research points to a delay and dysfunction in the development of the prefrontal cortex and disturbances in neurotransmission. According to studies of twins, the disorder has a strong genetic component. There are several candidate genes that may contribute to the disorder, but no definitive links have been discovered. Environmental factors, including exposure to certain pesticides, may also contribute to the development of ADHD in some patients. Treatment for ADHD often involves behavioral therapies and the prescription of stimulant medications, which paradoxically cause a calming effect in these patients.

Many people with ADHD have one or more other neurological disorders. (credit “chart design and illustration”: modification of work by Leigh Coriale; credit “data”: Drs. Biederman and Faraone, Massachusetts General Hospital).

Note:
Career Connection

Neurologist

Neurologists are physicians who specialize in disorders of the nervous system. They diagnose and treat disorders such as epilepsy, stroke, dementia, nervous system injuries, Parkinson’s disease, sleep disorders, and multiple sclerosis. Neurologists are medical doctors who have attended college, medical school, and completed three to four years of neurology residency.

When examining a new patient, a neurologist takes a full medical history and performs a complete physical exam. The physical exam contains specific tasks that are used to determine what areas of the brain, spinal cord, or peripheral nervous system may be damaged. For example, to check whether the hypoglossal nerve is functioning correctly, the neurologist will ask the patient to move his or her tongue in different ways. If the patient does not have full control over tongue movements, then the hypoglossal nerve may be damaged or there may be a lesion in the brainstem where the cell bodies of these neurons reside (or there could be damage to the tongue muscle itself).

Neurologists have other tools besides a physical exam they can use to diagnose particular problems in the nervous system. If the patient has had a seizure, for example, the neurologist can use electroencephalography (EEG), which involves taping electrodes to the scalp to record brain activity, to try to determine which brain regions are involved in the seizure. In suspected stroke patients, a neurologist can use a computerized tomography (CT) scan, which is a type of X-ray, to look for bleeding in the brain or a possible brain tumor. To treat patients with neurological problems, neurologists can prescribe medications or refer the patient to a neurosurgeon for surgery.

Note:

Link to Learning

This website allows you to see the different tests a neurologist might use to see what regions of the nervous system may be damaged in a patient.

Mental Illnesses

Mental illnesses are nervous system disorders that result in problems with thinking, mood, or relating with other people. These disorders are severe enough to affect a person's quality of life and often make it difficult for people to perform the routine tasks of daily living. Debilitating mental disorders plague approximately 12.5 million Americans (about 1 in 17 people) at an annual cost of more than $300 billion. There are several types of mental disorders including schizophrenia, major depression, bipolar disorder, anxiety disorders and phobias, post-traumatic stress disorders, and obsessive-compulsive disorder (OCD), among others. The American Psychiatric Association publishes the Diagnostic and Statistical Manual of Mental Disorders (or DSM), which describes the symptoms required for a patient to be diagnosed with a particular mental disorder. Each newly released version of the DSM contains different symptoms and classifications as scientists learn more about these disorders, their causes, and how they relate to each other. A more detailed discussion of two mental illnesses—schizophrenia and major depression—is given below.

Schizophrenia

Schizophrenia is a serious and often debilitating mental illness affecting one percent of people in the United States. Symptoms of the disease include the inability to differentiate between reality and imagination, inappropriate and unregulated emotional responses, difficulty thinking, and problems with

social situations. People with schizophrenia can suffer from hallucinations and hear voices; they may also suffer from delusions. Patients also have so-called “negative” symptoms like a flattened emotional state, loss of pleasure, and loss of basic drives. Many schizophrenic patients are diagnosed in their late adolescence or early 20s. The development of schizophrenia is thought to involve malfunctioning dopaminergic neurons and may also involve problems with glutamate signaling. Treatment for the disease usually requires antipsychotic medications that work by blocking dopamine receptors and decreasing dopamine neurotransmission in the brain. This decrease in dopamine can cause Parkinson’s disease-like symptoms in some patients. While some classes of antipsychotics can be quite effective at treating the disease, they are not a cure, and most patients must remain medicated for the rest of their lives.

Depression

Major depression affects approximately 6.7 percent of the adults in the United States each year and is one of the most common mental disorders. To be diagnosed with major depressive disorder, a person must have experienced a severely depressed mood lasting longer than two weeks along with other symptoms including a loss of enjoyment in activities that were previously enjoyed, changes in appetite and sleep schedules, difficulty concentrating, feelings of worthlessness, and suicidal thoughts. The exact causes of major depression are unknown and likely include both genetic and environmental risk factors. Some research supports the “classic monoamine hypothesis,” which suggests that depression is caused by a decrease in norepinephrine and serotonin neurotransmission. One argument against this hypothesis is the fact that some antidepressant medications cause an increase in norepinephrine and serotonin release within a few hours of beginning treatment—but clinical results of these medications are not seen until weeks later. This has led to alternative hypotheses: for example, dopamine may also be decreased in depressed patients, or it may actually be an increase in norepinephrine and serotonin that causes the disease, and antidepressants force a feedback loop that decreases this release. Treatments for depression include psychotherapy, electroconvulsive therapy, deep-brain stimulation, and prescription medications. There are several classes of

antidepressant medications that work through different mechanisms. For example, monoamine oxidase inhibitors (MAO inhibitors) block the enzyme that degrades many neurotransmitters (including dopamine, serotonin, norepinephrine), resulting in increased neurotransmitter in the synaptic cleft. Selective serotonin reuptake inhibitors (SSRIs) block the reuptake of serotonin into the presynaptic neuron. This blockage results in an increase in serotonin in the synaptic cleft. Other types of drugs such as norepinephrine-dopamine reuptake inhibitors and norepinephrine-serotonin reuptake inhibitors are also used to treat depression.

Other Neurological Disorders

There are several other neurological disorders that cannot be easily placed in the above categories. These include chronic pain conditions, cancers of the nervous system, epilepsy disorders, and stroke. Epilepsy and stroke are discussed below.

Epilepsy

Estimates suggest that up to three percent of people in the United States will be diagnosed with **epilepsy** in their lifetime. While there are several different types of epilepsy, all are characterized by recurrent seizures. Epilepsy itself can be a symptom of a brain injury, disease, or other illness. For example, people who have intellectual disability or ASD can experience seizures, presumably because the developmental wiring malfunctions that caused their disorders also put them at risk for epilepsy. For many patients, however, the cause of their epilepsy is never identified and is likely to be a combination of genetic and environmental factors. Often, seizures can be controlled with anticonvulsant medications. However, for very severe cases, patients may undergo brain surgery to remove the brain area where seizures originate.

Stroke

A stroke results when blood fails to reach a portion of the brain for a long enough time to cause damage. Without the oxygen supplied by blood flow, neurons in this brain region die. This neuronal death can cause many different symptoms—depending on the brain area affected— including headache, muscle weakness or paralysis, speech disturbances, sensory problems, memory loss, and confusion. Stroke is often caused by blood clots and can also be caused by the bursting of a weak blood vessel. Strokes are extremely common and are the third most common cause of death in the United States. On average one person experiences a stroke every 40 seconds in the United States. Approximately 75 percent of strokes occur in people older than 65. Risk factors for stroke include high blood pressure, diabetes, high cholesterol, and a family history of stroke. Smoking doubles the risk of stroke. Because a stroke is a medical emergency, patients with symptoms of a stroke should immediately go to the emergency room, where they can receive drugs that will dissolve any clot that may have formed. These drugs will not work if the stroke was caused by a burst blood vessel or if the stroke occurred more than three hours before arriving at the hospital. Treatment following a stroke can include blood pressure medication (to prevent future strokes) and (sometimes intense) physical therapy.

Section Summary

Some general themes emerge from the sampling of nervous system disorders presented above. The causes for most disorders are not fully understood—at least not for all patients—and likely involve a combination of nature (genetic mutations that become risk factors) and nurture (emotional trauma, stress, hazardous chemical exposure). Because the causes have yet to be fully determined, treatment options are often lacking and only address symptoms.

Review Questions

Exercise:

Problem:

Parkinson’s disease is a caused by the degeneration of neurons that release _________.

a. serotonin
b. dopamine
c. glutamate
d. norepinephrine

Solution:

B

Exercise:

Problem:

_________ medications are often used to treat patients with ADHD.

a. Tranquilizer
b. Antibiotic
c. Stimulant
d. Anti-seizure

Solution:

C

Exercise:

Problem: Strokes are often caused by _________.

a. neurodegeneration
b. blood clots or burst blood vessels
c. seizures
d. viruses

Solution:

B

Free Response

Exercise:

Problem: What are the main symptoms of Alzheimer’s disease?

Solution:

Symptoms of Alzheimer’s disease include disruptive memory loss, confusion about time or place, difficulties planning or executing tasks, poor judgment, and personality changes.

Exercise:

Problem:

What are possible treatments for patients with major depression?

Solution:

Possible treatments for patients with major depression include psychotherapy and prescription medications. MAO inhibitor drugs inhibit the breakdown of certain neurotransmitters (including dopamine, serotonin, norepinephrine) in the synaptic cleft. SSRI medications inhibit the reuptake of serotonin into the presynaptic neuron.

Glossary

Alzheimer’s disease
: neurodegenerative disorder characterized by problems with memory and thinking

attention deficit hyperactivity disorder (ADHD)
: neurodevelopmental disorder characterized by difficulty maintaining attention and controlling impulses

autism spectrum disorder (ASD)
: neurodevelopmental disorder characterized by impaired social interaction and communication abilities

epilepsy
: neurological disorder characterized by recurrent seizures

major depression
: mental illness characterized by prolonged periods of sadness

neurodegenerative disorder
: nervous system disorder characterized by the progressive loss of neurological functioning, usually caused by neuron death

Parkinson’s disease
: neurodegenerative disorder that affects the control of movement

schizophrenia
: mental disorder characterized by the inability to accurately perceive reality; patients often have difficulty thinking clearly and can suffer from delusions

Sensory Processes

By the end of this section, you will be able to:

- Identify the general and special senses in humans
- Describe three important steps in sensory perception
- Explain the concept of just-noticeable difference in sensory perception

Senses provide information about the body and its environment. Humans have five special senses: olfaction (smell), gustation (taste), equilibrium (balance and body position), vision, and hearing. Additionally, we possess general senses, also called somatosensation, which respond to stimuli like temperature, pain, pressure, and vibration. **Vestibular sensation**, which is an organism’s sense of spatial orientation and balance, **proprioception** (position of bones, joints, and muscles), and the sense of limb position that is used to track **kinesthesia** (limb movement) are part of somatosensation. Although the sensory systems associated with these senses are very different, all share a common function: to convert a stimulus (such as light, or sound, or the position of the body) into an electrical signal in the nervous system. This process is called **sensory transduction**.

There are two broad types of cellular systems that perform sensory transduction. In one, a neuron works with a **sensory receptor**, a cell, or cell process that is specialized to engage with and detect a specific stimulus. Stimulation of the sensory receptor activates the associated afferent neuron, which carries information about the stimulus to the central nervous system. In the second type of sensory transduction, a sensory nerve ending responds to a stimulus in the internal or external environment: this neuron constitutes the sensory receptor. Free nerve endings can be stimulated by several different stimuli, thus showing little receptor specificity. For example, pain receptors in your gums and teeth may be stimulated by temperature changes, chemical stimulation, or pressure.

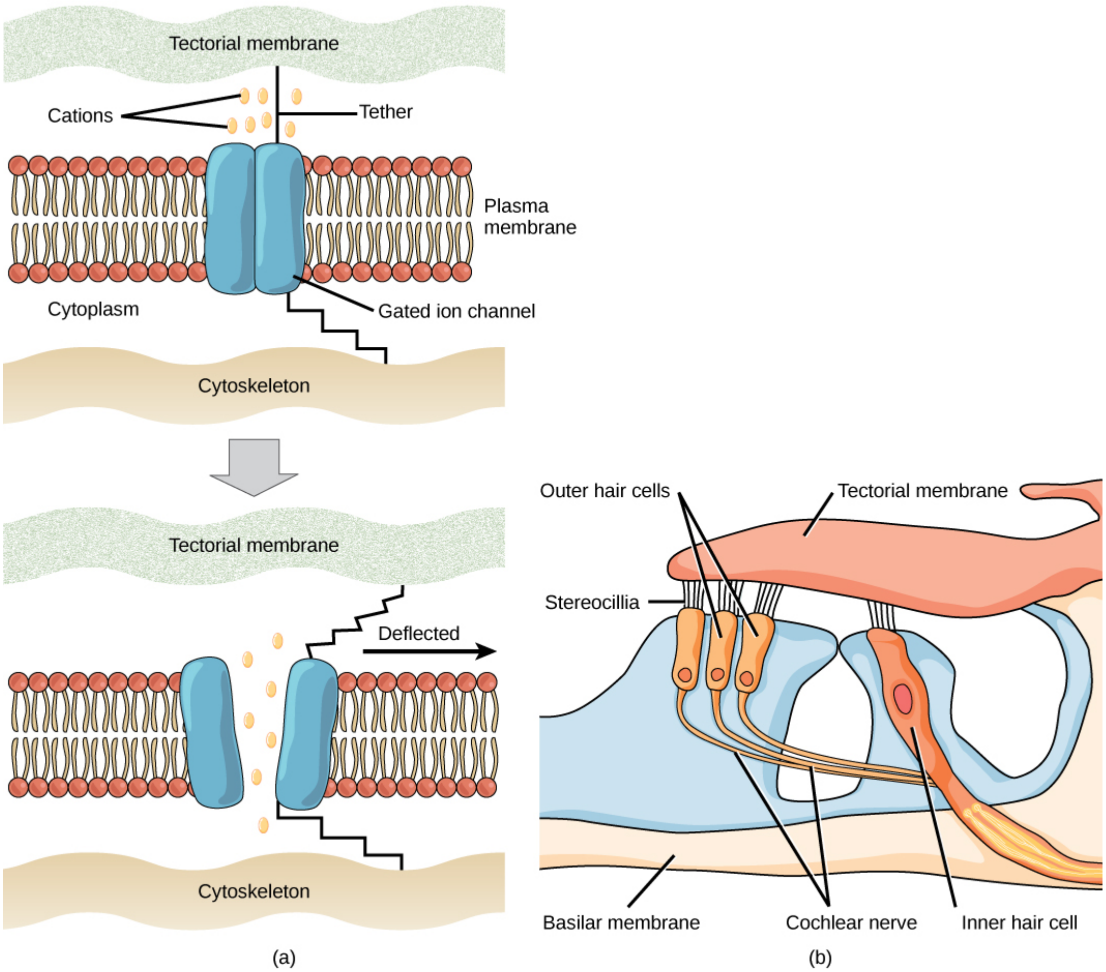
Reception

The first step in sensation is **reception**, which is the activation of sensory receptors by stimuli such as mechanical stimuli (being bent or squished, for example), chemicals, or temperature. The receptor can then respond to the

stimuli. The region in space in which a given sensory receptor can respond to a stimulus, be it far away or in contact with the body, is that receptor’s **receptive field**. Think for a moment about the differences in receptive fields for the different senses. For the sense of touch, a stimulus must come into contact with body. For the sense of hearing, a stimulus can be a moderate distance away (some baleen whale sounds can propagate for many kilometers). For vision, a stimulus can be very far away; for example, the visual system perceives light from stars at enormous distances.

Transduction

The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system. This takes place at the sensory receptor, and the change in electrical potential that is produced is called the **receptor potential**. How is sensory input, such as pressure on the skin, changed to a receptor potential? In this example, a type of receptor called a **mechanoreceptor** (as shown in [link]) possesses specialized membranes that respond to pressure. Disturbance of these dendrites by compressing them or bending them opens gated ion channels in the plasma membrane of the sensory neuron, changing its electrical potential. Recall that in the nervous system, a positive change of a neuron’s electrical potential (also called the membrane potential), depolarizes the neuron. Receptor potentials are graded potentials: the magnitude of these graded (receptor) potentials varies with the strength of the stimulus. If the magnitude of depolarization is sufficient (that is, if membrane potential reaches a threshold), the neuron will fire an action potential. In most cases, the correct stimulus impinging on a sensory receptor will drive membrane potential in a positive direction, although for some receptors, such as those in the visual system, this is not always the case.



(a) Mechanosensitive ion channels are gated ion channels that respond to mechanical deformation of the plasma membrane. A mechanosensitive channel is connected to the plasma membrane and the cytoskeleton by hair-like tethers. When pressure causes the extracellular matrix to move, the channel opens, allowing ions to enter or exit the cell. (b) Stereocilia in the human ear are connected to mechanosensitive ion channels. When a sound causes the stereocilia to move, mechanosensitive ion channels transduce the signal to the cochlear nerve.

Sensory receptors for different senses are very different from each other, and they are specialized according to the type of stimulus they sense: they

have receptor specificity. For example, touch receptors, light receptors, and sound receptors are each activated by different stimuli. Touch receptors are not sensitive to light or sound; they are sensitive only to touch or pressure. However, stimuli may be combined at higher levels in the brain, as happens with olfaction, contributing to our sense of taste.

Encoding and Transmission of Sensory Information

Four aspects of sensory information are encoded by sensory systems: the type of stimulus, the location of the stimulus in the receptive field, the duration of the stimulus, and the relative intensity of the stimulus. Thus, action potentials transmitted over a sensory receptor’s afferent axons encode one type of stimulus, and this segregation of the senses is preserved in other sensory circuits. For example, auditory receptors transmit signals over their own dedicated system, and electrical activity in the axons of the auditory receptors will be interpreted by the brain as an auditory stimulus—a sound.

The intensity of a stimulus is often encoded in the rate of action potentials produced by the sensory receptor. Thus, an intense stimulus will produce a more rapid train of action potentials, and reducing the stimulus will likewise slow the rate of production of action potentials. A second way in which intensity is encoded is by the number of receptors activated. An intense stimulus might initiate action potentials in a large number of adjacent receptors, while a less intense stimulus might stimulate fewer receptors. Integration of sensory information begins as soon as the information is received in the CNS, and the brain will further process incoming signals.

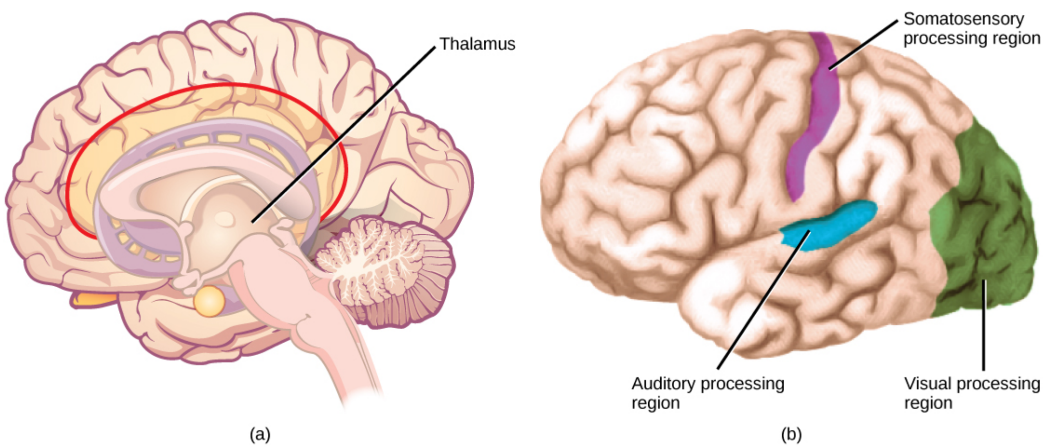
Perception

Perception is an individual’s interpretation of a sensation. Although perception relies on the activation of sensory receptors, perception happens not at the level of the sensory receptor, but at higher levels in the nervous system, in the brain. The brain distinguishes sensory stimuli through a

sensory pathway: action potentials from sensory receptors travel along neurons that are dedicated to a particular stimulus. These neurons are dedicated to that particular stimulus and synapse with particular neurons in the brain or spinal cord.

All sensory signals, except those from the olfactory system, are transmitted though the central nervous system and are routed to the thalamus and to the appropriate region of the cortex. Recall that the thalamus is a structure in the forebrain that serves as a clearinghouse and relay station for sensory (as well as motor) signals. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex ([link]) dedicated to processing that particular sense.

How are neural signals interpreted? Interpretation of sensory signals between individuals of the same species is largely similar, owing to the inherited similarity of their nervous systems; however, there are some individual differences. A good example of this is individual tolerances to a painful stimulus, such as dental pain, which certainly differ.



In humans, with the exception of olfaction, all sensory signals are routed from the (a) thalamus to (b) final processing regions in the cortex of the brain. (credit b: modification of work by Polina Tishina)

Note:
Scientific Method Connection

Just-Noticeable Difference

It is easy to differentiate between a one-pound bag of rice and a two-pound bag of rice. There is a one-pound difference, and one bag is twice as heavy as the other. However, would it be as easy to differentiate between a 20- and a 21-pound bag?

Question: What is the smallest detectible weight difference between a one-pound bag of rice and a larger bag? What is the smallest detectible difference between a 20-pound bag and a larger bag? In both cases, at what weights are the differences detected? This smallest detectible difference in stimuli is known as the just-noticeable difference (JND).

Background: Research background literature on JND and on Weber’s Law, a description of a proposed mathematical relationship between the overall magnitude of the stimulus and the JND. You will be testing JND of different weights of rice in bags. Choose a convenient increment that is to be stepped through while testing. For example, you could choose 10 percent increments between one and two pounds (1.1, 1.2, 1.3, 1.4, and so on) or 20 percent increments (1.2, 1.4, 1.6, and 1.8).

Hypothesis: Develop a hypothesis about JND in terms of percentage of the whole weight being tested (such as “the JND between the two small bags and between the two large bags is proportionally the same,” or “. . . is not proportionally the same.”) So, for the first hypothesis, if the JND between the one-pound bag and a larger bag is 0.2 pounds (that is, 20 percent; 1.0 pound feels the same as 1.1 pounds, but 1.0 pound feels less than 1.2 pounds), then the JND between the 20-pound bag and a larger bag will also be 20 percent. (So, 20 pounds feels the same as 22 pounds or 23 pounds, but 20 pounds feels less than 24 pounds.)

Test the hypothesis: Enlist 24 participants, and split them into two groups of 12. To set up the demonstration, assuming a 10 percent increment was selected, have the first group be the one-pound group. As a counter-balancing measure against a systematic error, however, six of the first group will compare one pound to two pounds, and step down in weight (1.0 to 2.0, 1.0 to 1.9, and so on.), while the other six will step up (1.0 to 1.1, 1.0 to 1.2, and so on). Apply the same principle to the 20-pound group (20 to 40, 20 to 38, and so on, and 20 to 22, 20 to 24, and so on). Given the large difference between 20 and 40 pounds, you may wish to use 30

pounds as your larger weight. In any case, use two weights that are easily detectable as different.

Record the observations: Record the data in a table similar to the table below. For the one-pound and 20-pound groups (base weights) record a plus sign (+) for each participant that detects a difference between the base weight and the step weight. Record a minus sign (-) for each participant that finds no difference. If one-tenth steps were not used, then replace the steps in the "Step Weight" columns with the step you are using.

Results of JND Testing (+ = difference; – = no difference)			
Step Weight	**One pound**	**20 pounds**	**Step Weight**
1.1			22
1.2			24
1.3			26
1.4			28
1.5			30
1.6			32
1.7			34
1.8			36
1.9			38
2.0			40

Analyze the data/report the results: What step weight did all participants find to be equal with one-pound base weight? What about the 20-pound group?
Draw a conclusion: Did the data support the hypothesis? Are the final weights proportionally the same? If not, why not? Do the findings adhere to Weber's Law? Weber's Law states that the concept that a just-noticeable difference in a stimulus is proportional to the magnitude of the original stimulus.

Section Summary

A sensory activation occurs when a physical or chemical stimulus is processed into a neural signal (sensory transduction) by a sensory receptor. Perception is an individual interpretation of a sensation and is a brain function. Humans have special senses: olfaction, gustation, equilibrium, and hearing, plus the general senses of somatosensation.

Sensory receptors are either specialized cells associated with sensory neurons or the specialized ends of sensory neurons that are a part of the peripheral nervous system, and they are used to receive information about the environment (internal or external). Each sensory receptor is modified for the type of stimulus it detects. For example, neither gustatory receptors nor auditory receptors are sensitive to light. Each sensory receptor is responsive to stimuli within a specific region in space, which is known as that receptor's receptive field. The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system.

All sensory signals, except those from the olfactory system, enter the central nervous system and are routed to the thalamus. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex dedicated to processing that particular sense.

Review Questions

Exercise:

Problem: Where does perception occur?

a. spinal cord
b. cerebral cortex
c. receptors
d. thalamus

Solution:

B

Exercise:

Problem:

If a person's cold receptors no longer convert cold stimuli into sensory signals, that person has a problem with the process of _________.

a. reception
b. transmission
c. perception
d. transduction

Solution:

D

Exercise:

Problem:

After somatosensory transduction, the sensory signal travels through the brain as a(n) ______ signal.

a. electrical
b. pressure

c. optical
d. thermal

Solution:

A

Free Response

Exercise:

Problem:

If a person sustains damage to axons leading from sensory receptors to the central nervous system, which step or steps of sensory perception will be affected?

Solution:

Transmission of sensory information from the receptor to the central nervous system will be impaired, and thus, perception of stimuli, which occurs in the brain, will be halted.

Exercise:

Problem:

In what way does the overall magnitude of a stimulus affect the just-noticeable difference in the perception of that stimulus?

Solution:

The just-noticeable difference is a fraction of the overall magnitude of the stimulus and seems to be a relatively fixed proportion (such as 10 percent) whether the stimulus is large (such as a very heavy object) or small (such as a very light object).

Glossary

kinesthesia
: sense of body movement

mechanoreceptor
: sensory receptor modified to respond to mechanical disturbance such as being bent, touch, pressure, motion, and sound

perception
: individual interpretation of a sensation; a brain function

proprioception
: sense of limb position; used to track kinesthesia

reception
: receipt of a signal (such as light or sound) by sensory receptors

receptive field
: region in space in which a stimulus can activate a given sensory receptor

receptor potential
: membrane potential in a sensory receptor in response to detection of a stimulus

sensory receptor
: specialized neuron or other cells associated with a neuron that is modified to receive specific sensory input

sensory transduction
: conversion of a sensory stimulus into electrical energy in the nervous system by a change in the membrane potential

vestibular sense
: sense of spatial orientation and balance

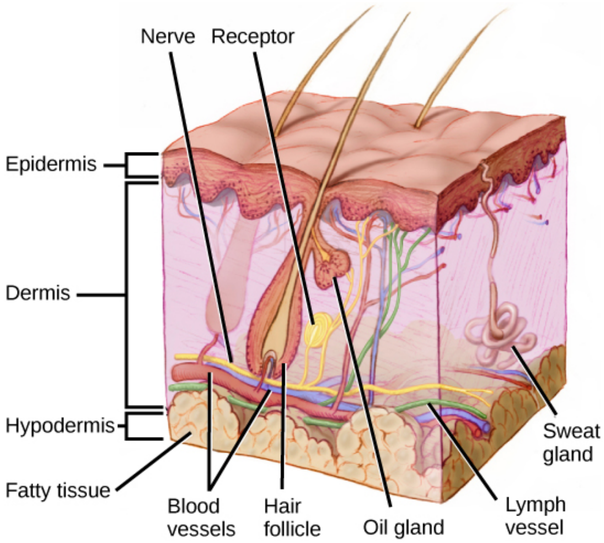
Somatosensation

By the end of this section, you will be able to:

- Describe four important mechanoreceptors in human skin
- Describe the topographical distribution of somatosensory receptors between glabrous and hairy skin
- Explain why the perception of pain is subjective

Somatosensation is a mixed sensory category and includes all sensation received from the skin and mucous membranes, as well from as the limbs and joints. Somatosensation is also known as tactile sense, or more familiarly, as the sense of touch. Somatosensation occurs all over the exterior of the body and at some interior locations as well. A variety of receptor types—embedded in the skin, mucous membranes, muscles, joints, internal organs, and cardiovascular system—play a role.

Recall that the epidermis is the outermost layer of skin in mammals. It is relatively thin, is composed of keratin-filled cells, and has no blood supply. The epidermis serves as a barrier to water and to invasion by pathogens. Below this, the much thicker dermis contains blood vessels, sweat glands, hair follicles, lymph vessels, and lipid-secreting sebaceous glands ([link]). Below the epidermis and dermis is the subcutaneous tissue, or hypodermis, the fatty layer that contains blood vessels, connective tissue, and the axons of sensory neurons. The hypodermis, which holds about 50 percent of the body's fat, attaches the dermis to the bone and muscle, and supplies nerves and blood vessels to the dermis.



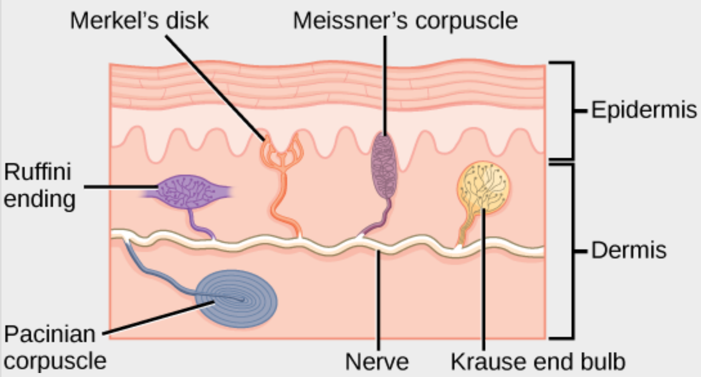
Mammalian skin has three layers: an epidermis, a dermis, and a hypodermis. (credit: modification of work by Don Bliss, National Cancer Institute)

Somatosensory Receptors

Sensory receptors are classified into five categories: mechanoreceptors, thermoreceptors, proprioceptors, pain receptors, and chemoreceptors. These categories are based on the nature of stimuli each receptor class transduces. What is commonly referred to as "touch" involves more than one kind of stimulus and more than one kind of receptor. Mechanoreceptors in the skin are described as encapsulated (that is, surrounded by a capsule) or unencapsulated (a group that includes free nerve endings). A **free nerve ending**, as its name implies, is an unencapsulated dendrite of a sensory neuron. Free nerve endings are the most common nerve endings in skin, and they extend into the middle of the epidermis. Free nerve endings are sensitive to painful stimuli, to hot and cold, and to light touch. They are slow to adjust to a stimulus and so are less sensitive to abrupt changes in stimulation.

There are three classes of mechanoreceptors: tactile, proprioceptors, and baroreceptors. Mechanoreceptors sense stimuli due to physical deformation of their plasma membranes. They contain mechanically gated ion channels whose gates open or close in response to pressure, touch, stretching, and sound." There are four primary tactile mechanoreceptors in human skin: Merkel's disks, Meissner's corpuscles, Ruffini endings, and Pacinian corpuscle; two are located toward the surface of the skin and two are located deeper. A fifth type of mechanoreceptor, Krause end bulbs, are found only in specialized regions. **Merkel's disks** (shown in [link]) are found in the upper layers of skin near the base of the epidermis, both in skin that has hair and on **glabrous** skin, that is, the hairless skin found on the palms and fingers, the soles of the feet, and the lips of humans and other primates. Merkel's disks are densely distributed in the fingertips and lips. They are slow-adapting, encapsulated nerve endings, and they respond to light touch. Light touch, also known as discriminative touch, is a light pressure that allows the location of a stimulus to be pinpointed. The receptive fields of Merkel's disks are small with well-defined borders. That makes them finely sensitive to edges and they come into use in tasks such as typing on a keyboard.

Note:
Art Connection



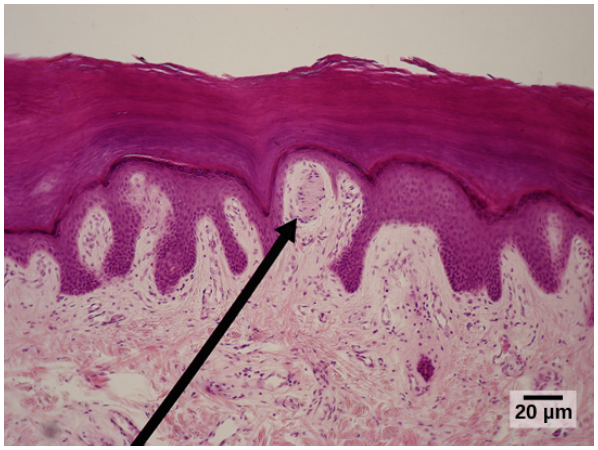
Four of the primary mechanoreceptors in human skin are shown. Merkel's disks, which are unencapsulated, respond to light touch. Meissner's

corpuscles, Ruffini endings, Pacinian corpuscles, and Krause end bulbs are all encapsulated. Meissner's corpuscles respond to touch and low-frequency vibration. Ruffini endings detect stretch, deformation within joints, and warmth. Pacinian corpuscles detect transient pressure and high-frequency vibration. Krause end bulbs detect cold.

Which of the following statements about mechanoreceptors is false?

a. Pacini corpuscles are found in both glabrous and hairy skin.
b. Merkel's disks are abundant on the fingertips and lips.
c. Ruffini endings are encapsulated mechanoreceptors.
d. Meissner's corpuscles extend into the lower dermis.

Meissner's corpuscles, (shown in [link]) also known as tactile corpuscles, are found in the upper dermis, but they project into the epidermis. They, too, are found primarily in the glabrous skin on the fingertips and eyelids. They respond to fine touch and pressure, but they also respond to low-frequency vibration or flutter. They are rapidly adapting, fluid-filled, encapsulated neurons with small, well-defined borders and are responsive to fine details. Like Merkel's disks, Meissner's corpuscles are not as plentiful in the palms as they are in the fingertips.

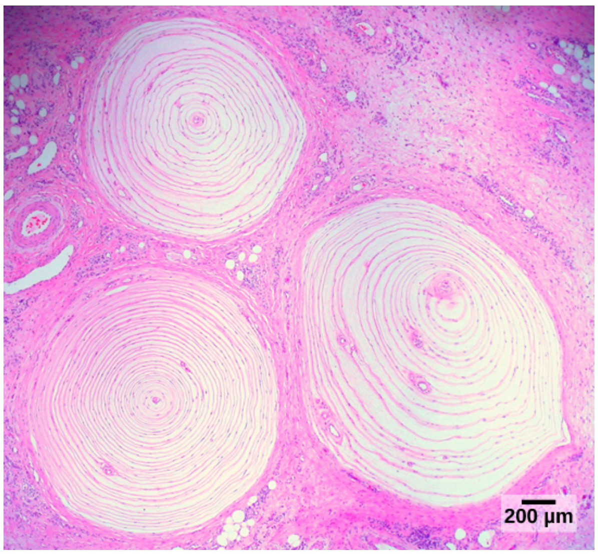


Meissner corpuscles in the fingertips, such as the one viewed here using bright field light microscopy, allow for touch discrimination of fine detail. (credit: modification of work by "Wbensmith"/Wikimedia Commons; scale-bar data from Matt Russell)

Deeper in the epidermis, near the base, are **Ruffini endings**, which are also known as bulbous corpuscles. They are found in both glabrous and hairy skin. These are slow-adapting, encapsulated mechanoreceptors that detect skin stretch and deformations within joints, so they provide valuable feedback for gripping objects and controlling finger position and movement. Thus, they also contribute to proprioception and kinesthesia. Ruffini endings also detect warmth. Note that these warmth detectors are situated deeper in the skin than are the cold detectors. It is not surprising, then, that humans detect cold stimuli before they detect warm stimuli.

Pacinian corpuscles (seen in [link]) are located deep in the dermis of both glabrous and hairy skin and are structurally similar to Meissner’s corpuscles; they are found in the bone periosteum, joint capsules, pancreas

and other viscera, breast, and genitals. They are rapidly adapting mechanoreceptors that sense deep transient (but not prolonged) pressure and high-frequency vibration. Pacinian receptors detect pressure and vibration by being compressed, stimulating their internal dendrites. There are fewer Pacinian corpuscles and Ruffini endings in skin than there are Merkel’s disks and Meissner’s corpuscles.



Pacinian corpuscles, such as these visualized using bright field light microscopy, detect pressure (touch) and high-frequency vibration. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

In proprioception, proprioceptive and kinesthetic signals travel through myelinated afferent neurons running from the spinal cord to the medulla. Neurons are not physically connected, but communicate via neurotransmitters secreted into synapses or “gaps” between communicating

neurons. Once in the medulla, the neurons continue carrying the signals to the thalamus.

Muscle spindles are stretch receptors that detect the amount of stretch, or lengthening of muscles. Related to these are **Golgi tendon organs**, which are tension receptors that detect the force of muscle contraction. Proprioceptive and kinesthetic signals come from limbs. Unconscious proprioceptive signals run from the spinal cord to the cerebellum, the brain region that coordinates muscle contraction, rather than to the thalamus, like most other sensory information.

Barorecptors detect pressure changes in an organ. They are found in the walls of the carotid artery and the aorta where they monitor blood pressure, and in the lungs where they detect the degree of lung expansion. Stretch receptors are found at various sites in the digestive and urinary systems.

In addition to these two types of deeper receptors, there are also rapidly adapting hair receptors, which are found on nerve endings that wrap around the base of hair follicles. There are a few types of hair receptors that detect slow and rapid hair movement, and they differ in their sensitivity to movement. Some hair receptors also detect skin deflection, and certain rapidly adapting hair receptors allow detection of stimuli that have not yet touched the skin.

Integration of Signals from Mechanoreceptors

The configuration of the different types of receptors working in concert in human skin results in a very refined sense of touch. The nociceptive receptors—those that detect pain—are located near the surface. Small, finely calibrated mechanoreceptors—Merkel’s disks and Meissner’s corpuscles—are located in the upper layers and can precisely localize even gentle touch. The large mechanoreceptors—Pacinian corpuscles and Ruffini endings—are located in the lower layers and respond to deeper touch. (Consider that the deep pressure that reaches those deeper receptors would not need to be finely localized.) Both the upper and lower layers of the skin hold rapidly and slowly adapting receptors. Both primary somatosensory

cortex and secondary cortical areas are responsible for processing the complex picture of stimuli transmitted from the interplay of mechanoreceptors.

Density of Mechanoreceptors

The distribution of touch receptors in human skin is not consistent over the body. In humans, touch receptors are less dense in skin covered with any type of hair, such as the arms, legs, torso, and face. Touch receptors are denser in glabrous skin (the type found on human fingertips and lips, for example), which is typically more sensitive and is thicker than hairy skin (4 to 5 mm versus 2 to 3 mm).

How is receptor density estimated in a human subject? The relative density of pressure receptors in different locations on the body can be demonstrated experimentally using a two-point discrimination test. In this demonstration, two sharp points, such as two thumbtacks, are brought into contact with the subject’s skin (though not hard enough to cause pain or break the skin). The subject reports if he or she feels one point or two points. If the two points are felt as one point, it can be inferred that the two points are both in the receptive field of a single sensory receptor. If two points are felt as two separate points, each is in the receptive field of two separate sensory receptors. The points could then be moved closer and re-tested until the subject reports feeling only one point, and the size of the receptive field of a single receptor could be estimated from that distance.

Thermoreception

In addition to Krause end bulbs that detect cold and Ruffini endings that detect warmth, there are different types of cold receptors on some free nerve endings: thermoreceptors, located in the dermis, skeletal muscles, liver, and hypothalamus, that are activated by different temperatures. Their pathways into the brain run from the spinal cord through the thalamus to the primary somatosensory cortex. Warmth and cold information from the face travels through one of the cranial nerves to the brain. You know from

experience that a tolerably cold or hot stimulus can quickly progress to a much more intense stimulus that is no longer tolerable. Any stimulus that is too intense can be perceived as pain because temperature sensations are conducted along the same pathways that carry pain sensations

Pain

Pain is the name given to **nociception**, which is the neural processing of injurious stimuli in response to tissue damage. Pain is caused by true sources of injury, such as contact with a heat source that causes a thermal burn or contact with a corrosive chemical. But pain also can be caused by harmless stimuli that mimic the action of damaging stimuli, such as contact with capsaicins, the compounds that cause peppers to taste hot and which are used in self-defense pepper sprays and certain topical medications. Peppers taste "hot" because the protein receptors that bind capsaicin open the same calcium channels that are activated by warm receptors.

Nociception starts at the sensory receptors, but pain, inasmuch as it is the perception of nociception, does not start until it is communicated to the brain. There are several nociceptive pathways to and through the brain. Most axons carrying nociceptive information into the brain from the spinal cord project to the thalamus (as do other sensory neurons) and the neural signal undergoes final processing in the primary somatosensory cortex. Interestingly, one nociceptive pathway projects not to the thalamus but directly to the hypothalamus in the forebrain, which modulates the cardiovascular and neuroendocrine functions of the autonomic nervous system. Recall that threatening—or painful—stimuli stimulate the sympathetic branch of the visceral sensory system, readying a fight-or-flight response.

Note:
Link to Learning

View this video that animates the five phases of nociceptive pain. https://www.openstaxcollege.org/l/nociceptive

Section Summary

Somatosensation includes all sensation received from the skin and mucous membranes, as well as from the limbs and joints. Somatosensation occurs all over the exterior of the body and at some interior locations as well, and a variety of receptor types, embedded in the skin and mucous membranes, play a role.

There are several types of specialized sensory receptors. Rapidly adapting free nerve endings detect nociception, hot and cold, and light touch. Slowly adapting, encapsulated Merkel's disks are found in fingertips and lips, and respond to light touch. Meissner's corpuscles, found in glabrous skin, are rapidly adapting, encapsulated receptors that detect touch, low-frequency vibration, and flutter. Ruffini endings are slowly adapting, encapsulated receptors that detect skin stretch, joint activity, and warmth. Hair receptors are rapidly adapting nerve endings wrapped around the base of hair follicles that detect hair movement and skin deflection. Finally, Pacinian corpuscles are encapsulated, rapidly adapting receptors that detect transient pressure and high-frequency vibration.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about mechanoreceptors is false?

a. Pacini corpuscles are found in both glabrous and hairy skin.
b. Merkel's disks are abundant on the fingertips and lips.
c. Ruffini endings are encapsulated mechanoreceptors.
d. Meissner's corpuscles extend into the lower dermis.

Solution:

[link] D

Review Questions

Exercise:

Problem:

______ are found only in ______ skin, and detect skin deflection.

a. Meissner's corpuscles: hairy
b. Merkel's disks: glabrous
c. hair receptors: hairy
d. Krause end bulbs: hairy

Solution:

B

Exercise:

Problem:

If you were to burn your epidermis, what receptor type would you most likely burn?

a. free nerve endings
b. Ruffini endings
c. Pacinian corpuscle
d. hair receptors

Solution:

A

Free Response

Exercise:

Problem:

What can be inferred about the relative sizes of the areas of cortex that process signals from skin not densely innervated with sensory receptors and skin that is densely innervated with sensory receptors?

Solution:

The cortical areas serving skin that is densely innervated likely are larger than those serving skin that is less densely innervated.

Glossary

free nerve ending
: ending of an afferent neuron that lacks a specialized structure for detection of sensory stimuli; some respond to touch, pain, or temperature

glabrous
: describes the non-hairy skin found on palms and fingers, soles of feet, and lips of humans and other primates

Golgi tendon organ

muscular proprioceptive tension receptor that provides the sensory component of the Golgi tendon reflex

Meissner’s corpuscle
: (also, tactile corpuscle) encapsulated, rapidly-adapting mechanoreceptor in the skin that responds to light touch

Merkel's disc
: unencapsulated, slowly-adapting mechanoreceptor in the skin that responds to touch

muscle spindle
: proprioceptive stretch receptor that lies within a muscle and that shortens the muscle to an optimal length for efficient contraction

nociception
: neural processing of noxious (such as damaging) stimuli

Pacinian corpuscle
: encapsulated mechanoreceptor in the skin that responds to deep pressure and vibration

Ruffini ending
: (also, bulbous corpuscle) slowly-adapting mechanoreceptor in the skin that responds to skin stretch and joint position

Taste and Smell

By the end of this section, you will be able to:

- Explain in what way smell and taste stimuli differ from other sensory stimuli
- Identify the five primary tastes that can be distinguished by humans
- Explain in anatomical terms why a dog's sense of smell is more acute than a human's

Taste, also called **gustation**, and smell, also called **olfaction**, are the most interconnected senses in that both involve molecules of the stimulus entering the body and bonding to receptors. Smell lets an animal sense the presence of food or other animals—whether potential mates, predators, or prey—or other chemicals in the environment that can impact their survival. Similarly, the sense of taste allows animals to discriminate between types of foods. While the value of a sense of smell is obvious, what is the value of a sense of taste? Different tasting foods have different attributes, both helpful and harmful. For example, sweet-tasting substances tend to be highly caloric, which could be necessary for survival in lean times. Bitterness is associated with toxicity, and sourness is associated with spoiled food. Salty foods are valuable in maintaining homeostasis by helping the body retain water and by providing ions necessary for cells to function.

Tastes and Odors

Both taste and odor stimuli are molecules taken in from the environment. The primary tastes detected by humans are sweet, sour, bitter, salty and umami. The first four tastes need little explanation. The identification of **umami** as a fundamental taste occurred fairly recently—it was identified in 1908 by Japanese scientist Kikunae Ikeda while he worked with seaweed broth, but it was not widely accepted as a taste that could be physiologically distinguished until many years later. The taste of umami, also known as savoriness, is attributable to the taste of the amino acid L-glutamate. In fact, monosodium glutamate, or MSG, is often used in cooking to enhance the savory taste of certain foods. What is the adaptive value of being able to distinguish umami? Savory substances tend to be high in protein.

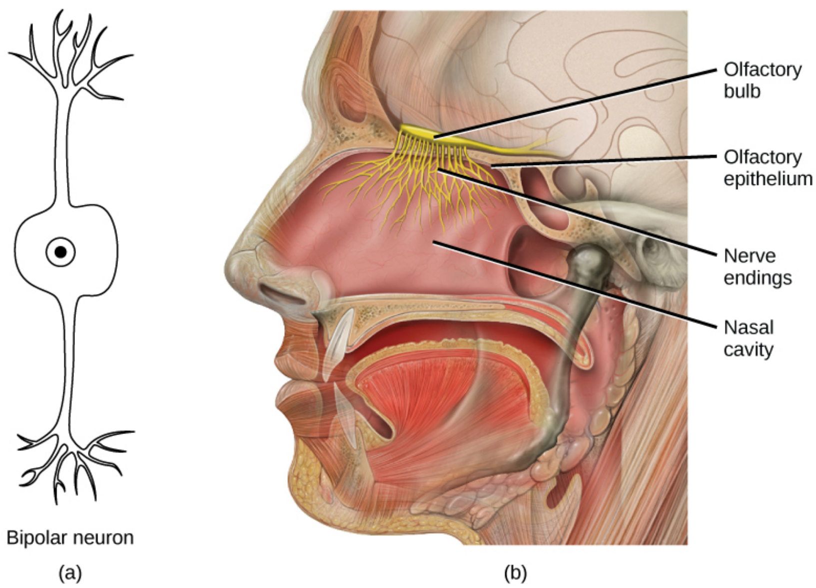
All odors that we perceive are molecules in the air we breathe. If a substance does not release molecules into the air from its surface, it has no smell. And if a human or other animal does not have a receptor that recognizes a specific molecule, then that molecule has no smell. Humans have about 350 olfactory receptor subtypes that work in various combinations to allow us to sense about 10,000 different odors. Compare that to mice, for example, which have about 1,300 olfactory receptor types, and therefore probably sense more odors. Both odors and tastes involve molecules that stimulate specific chemoreceptors. Although humans commonly distinguish taste as one sense and smell as another, they work together to create the perception of flavor. A person’s perception of flavor is reduced if he or she has congested nasal passages.

Reception and Transduction

Odorants (odor molecules) enter the nose and dissolve in the olfactory epithelium, the mucosa at the back of the nasal cavity (as illustrated in [link]). The **olfactory epithelium** is a collection of specialized olfactory receptors in the back of the nasal cavity that spans an area about 5 cm^2 in humans. Recall that sensory cells are neurons. An **olfactory receptor**, which is a dendrite of a specialized neuron, responds when it binds certain molecules inhaled from the environment by sending impulses directly to the olfactory bulb of the brain. Humans have about 12 million olfactory receptors, distributed among hundreds of different receptor types that respond to different odors. Twelve million seems like a large number of receptors, but compare that to other animals: rabbits have about 100 million, most dogs have about 1 billion, and bloodhounds—dogs selectively bred for their sense of smell—have about 4 billion. The overall size of the olfactory epithelium also differs between species, with that of bloodhounds, for example, being many times larger than that of humans.

Olfactory neurons are **bipolar neurons** (neurons with two processes from the cell body). Each neuron has a single dendrite buried in the olfactory epithelium, and extending from this dendrite are 5 to 20 receptor-laden, hair-like cilia that trap odorant molecules. The sensory receptors on the cilia are proteins, and it is the variations in their amino acid chains that make the receptors sensitive to different odorants. Each olfactory sensory neuron has

only one type of receptor on its cilia, and the receptors are specialized to detect specific odorants, so the bipolar neurons themselves are specialized. When an odorant binds with a receptor that recognizes it, the sensory neuron associated with the receptor is stimulated. Olfactory stimulation is the only sensory information that directly reaches the cerebral cortex, whereas other sensations are relayed through the thalamus.



In the human olfactory system, (a) bipolar olfactory neurons extend from (b) the olfactory epithelium, where olfactory receptors are located, to the olfactory bulb. (credit: modification of work by Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist)

Note:
Evolution Connection
Pheromones

A **pheromone** is a chemical released by an animal that affects the behavior or physiology of animals of the same species. Pheromonal signals can have profound effects on animals that inhale them, but pheromones apparently are not consciously perceived in the same way as other odors. There are several different types of pheromones, which are released in urine or as glandular secretions. Certain pheromones are attractants to potential mates, others are repellants to potential competitors of the same sex, and still others play roles in mother-infant attachment. Some pheromones can also influence the timing of puberty, modify reproductive cycles, and even prevent embryonic implantation. While the roles of pheromones in many nonhuman species are important, pheromones have become less important in human behavior over evolutionary time compared to their importance to organisms with more limited behavioral repertoires.

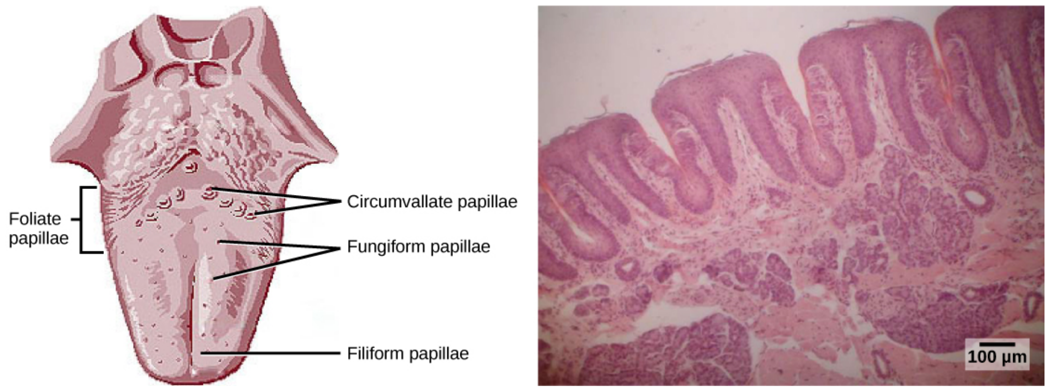
The vomeronasal organ (VNO, or Jacobson’s organ) is a tubular, fluid-filled, olfactory organ present in many vertebrate animals that sits adjacent to the nasal cavity. It is very sensitive to pheromones and is connected to the nasal cavity by a duct. When molecules dissolve in the mucosa of the nasal cavity, they then enter the VNO where the pheromone molecules among them bind with specialized pheromone receptors. Upon exposure to pheromones from their own species or others, many animals, including cats, may display the flehmen response (shown in [link]), a curling of the upper lip that helps pheromone molecules enter the VNO.

Pheromonal signals are sent, not to the main olfactory bulb, but to a different neural structure that projects directly to the amygdala (recall that the amygdala is a brain center important in emotional reactions, such as fear). The pheromonal signal then continues to areas of the hypothalamus that are key to reproductive physiology and behavior. While some scientists assert that the VNO is apparently functionally vestigial in humans, even though there is a similar structure located near human nasal cavities, others are researching it as a possible functional system that may, for example, contribute to synchronization of menstrual cycles in women living in close proximity.

The flehmen response in this tiger results in the curling of the upper lip and helps airborne pheromone molecules enter the vomeronasal organ. (credit: modification of work by "chadh"/Flickr)

Taste

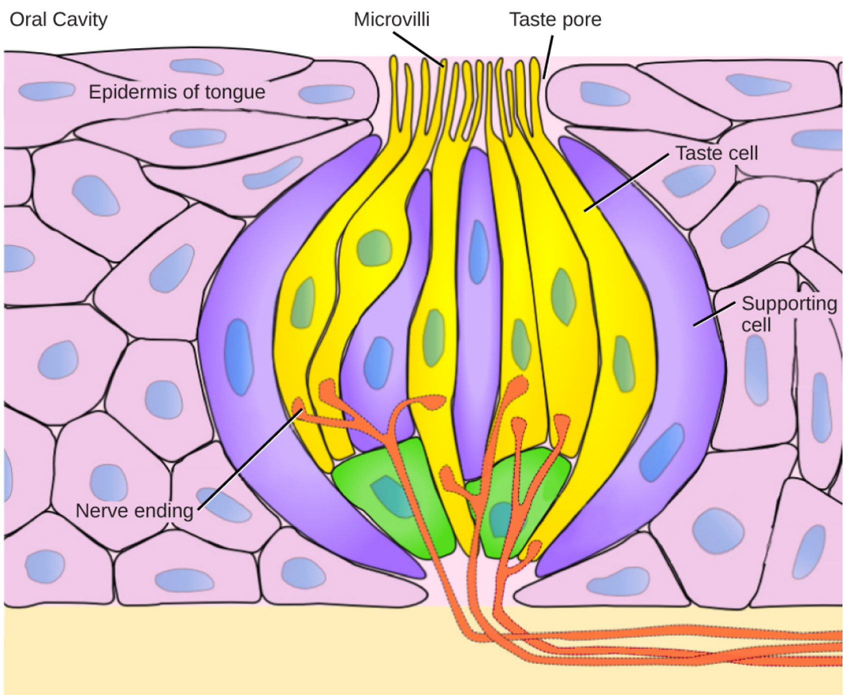
Detecting a taste (gustation) is fairly similar to detecting an odor (olfaction), given that both taste and smell rely on chemical receptors being stimulated by certain molecules. The primary organ of taste is the taste bud. A **taste bud** is a cluster of gustatory receptors (taste cells) that are located within the bumps on the tongue called **papillae** (singular: papilla) (illustrated in [link]). There are several structurally distinct papillae. Filiform papillae, which are located across the tongue, are tactile, providing friction that helps the tongue move substances, and contain no taste cells. In contrast, fungiform papillae, which are located mainly on the anterior two-thirds of the tongue, each contain one to eight taste buds and also have receptors for pressure and temperature. The large circumvallate papillae contain up to 100 taste buds and form a V near the posterior margin of the tongue.



(a) Foliate, circumvallate, and fungiform papillae are located on different regions of the tongue. (b) Foliate papillae are prominent protrusions on this light micrograph. (credit a: modification of work by NCI; scale-bar data from Matt Russell)

In addition to those two types of chemically and mechanically sensitive papillae are foliate papillae—leaf-like papillae located in parallel folds along the edges and toward the back of the tongue, as seen in the [link] micrograph. Foliate papillae contain about 1,300 taste buds within their folds. Finally, there are circumvallate papillae, which are wall-like papillae in the shape of an inverted “V” at the back of the tongue. Each of these papillae is surrounded by a groove and contains about 250 taste buds.

Each taste bud’s taste cells are replaced every 10 to 14 days. These are elongated cells with hair-like processes called microvilli at the tips that extend into the taste bud pore (illustrate in [link]). Food molecules (**tastants**) are dissolved in saliva, and they bind with and stimulate the receptors on the microvilli. The receptors for tastants are located across the outer portion and front of the tongue, outside of the middle area where the filiform papillae are most prominent.



Pores in the tongue allow tastants to enter taste pores in the tongue. (credit: modification of work by Vincenzo Rizzo)

In humans, there are five primary tastes, and each taste has only one corresponding type of receptor. Thus, like olfaction, each receptor is specific to its stimulus (tastant). Transduction of the five tastes happens through different mechanisms that reflect the molecular composition of the tastant. A salty tastant (containing NaCl) provides the sodium ions (Na^+) that enter the taste neurons and excite them directly. Sour tastants are acids and belong to the thermoreceptor protein family. Binding of an acid or other sour-tasting molecule triggers a change in the ion channel and these increase hydrogen ion (H^+) concentrations in the taste neurons, thus depolarizing them. Sweet, bitter, and umami tastants require a G-protein coupled receptor. These tastants bind to their respective receptors, thereby exciting the specialized neurons associated with them.

Both tasting abilities and sense of smell change with age. In humans, the senses decline dramatically by age 50 and continue to decline. A child may

find a food to be too spicy, whereas an elderly person may find the same food to be bland and unappetizing.

Note:
Link to Learning

View this animation that shows how the sense of taste works.

Smell and Taste in the Brain

Olfactory neurons project from the olfactory epithelium to the olfactory bulb as thin, unmyelinated axons. The **olfactory bulb** is composed of neural clusters called **glomeruli**, and each glomerulus receives signals from one type of olfactory receptor, so each glomerulus is specific to one odorant. From glomeruli, olfactory signals travel directly to the olfactory cortex and then to the frontal cortex and the thalamus. Recall that this is a different path from most other sensory information, which is sent directly to the thalamus before ending up in the cortex. Olfactory signals also travel directly to the amygdala, thereafter reaching the hypothalamus, thalamus, and frontal cortex. The last structure that olfactory signals directly travel to is a cortical center in the temporal lobe structure important in spatial, autobiographical, declarative, and episodic memories. Olfaction is finally processed by areas of the brain that deal with memory, emotions, reproduction, and thought.

Taste neurons project from taste cells in the tongue, esophagus, and palate to the medulla, in the brainstem. From the medulla, taste signals travel to the thalamus and then to the primary gustatory cortex. Information from

different regions of the tongue is segregated in the medulla, thalamus, and cortex.

Section Summary

There are five primary tastes in humans: sweet, sour, bitter, salty, and umami. Each taste has its own receptor type that responds only to that taste. Tastants enter the body and are dissolved in saliva. Taste cells are located within taste buds, which are found on three of the four types of papillae in the mouth.

Regarding olfaction, there are many thousands of odorants, but humans detect only about 10,000. Like taste receptors, olfactory receptors are each responsive to only one odorant. Odorants dissolve in nasal mucosa, where they excite their corresponding olfactory sensory cells. When these cells detect an odorant, they send their signals to the main olfactory bulb and then to other locations in the brain, including the olfactory cortex.

Review Questions

Exercise:

Problem: Which of the following has the fewest taste receptors?

a. fungiform papillae
b. circumvallate papillae
c. foliate papillae
d. filiform papillae

Solution:

D

Exercise:

Problem:

How many different taste molecules do taste cells each detect?

a. one
b. five
c. ten
d. It depends on the spot on the tongue

Solution:

A

Exercise:

Problem: Salty foods activate the taste cells by ______.

a. exciting the taste cell directly
b. causing hydrogen ions to enter the cell
c. causing sodium channels to close
d. binding directly to the receptors

Solution:

A

Exercise:

Problem:

All sensory signals except ______ travel to the ______ in the brain before the cerebral cortex.

a. vision; thalamus
b. olfaction; thalamus
c. vision; cranial nerves
d. olfaction; cranial nerves

Solution:

B

Free Response

Exercise:

Problem:

From the perspective of the recipient of the signal, in what ways do pheromones differ from other odorants?

Solution:

Pheromones may not be consciously perceived, and pheromones can have direct physiological and behavioral effects on their recipients.

Exercise:

Problem:

What might be the effect on an animal of not being able to perceive taste?

Solution:

The animal might not be able to recognize the differences in food sources and thus might not be able to discriminate between spoiled food and safe food or between foods that contain necessary nutrients, such as proteins, and foods that do not.

Glossary

bipolar neuron
: neuron with two processes from the cell body, typically in opposite directions

glomerulus
in the olfactory bulb, one of the two neural clusters that receives signals from one type of olfactory receptor

gustation
sense of taste

odorant
airborne molecule that stimulates an olfactory receptor

olfaction
sense of smell

olfactory bulb
neural structure in the vertebrate brain that receives signals from olfactory receptors

olfactory epithelium
specialized tissue in the nasal cavity where olfactory receptors are located

olfactory receptor
dendrite of a specialized neuron

papilla
one of the small bump-like projections from the tongue

pheromone
substance released by an animal that can affect the physiology or behavior of other animals

tastant
food molecule that stimulates gustatory receptors

taste bud
clusters of taste cells

umami

one of the five basic tastes, which is described as “savory” and which may be largely the taste of L-glutamate

Hearing and Vestibular Sensation
By the end of this section, you will be able to:

- Describe the relationship of amplitude and frequency of a sound wave to attributes of sound
- Trace the path of sound through the auditory system to the site of transduction of sound
- Identify the structures of the vestibular system that respond to gravity

Audition, or hearing, is important to humans and to other animals for many different interactions. It enables an organism to detect and receive information about danger, such as an approaching predator, and to participate in communal exchanges like those concerning territories or mating. On the other hand, although it is physically linked to the auditory system, the vestibular system is not involved in hearing. Instead, an animal’s vestibular system detects its own movement, both linear and angular acceleration and deceleration, and balance.

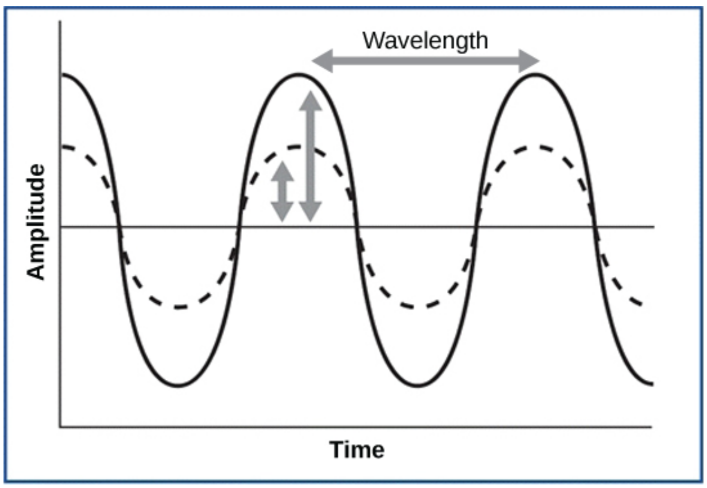
Sound

Auditory stimuli are sound waves, which are mechanical, pressure waves that move through a medium, such as air or water. There are no sound waves in a vacuum since there are no air molecules to move in waves. The speed of sound waves differs, based on altitude, temperature, and medium, but at sea level and a temperature of 20º C (68º F), sound waves travel in the air at about 343 meters per second.

As is true for all waves, there are four main characteristics of a sound wave: frequency, wavelength, period, and amplitude. Frequency is the number of waves per unit of time, and in sound is heard as pitch. High-frequency (≥15.000Hz) sounds are higher-pitched (short wavelength) than low-frequency (long wavelengths; ≤100Hz) sounds. Frequency is measured in cycles per second, and for sound, the most commonly used unit is hertz (Hz), or cycles per second. Most humans can perceive sounds with frequencies between 30 and 20,000 Hz. Women are typically better at hearing high frequencies, but everyone’s ability to hear high frequencies decreases with age. Dogs detect up to about 40,000 Hz; cats, 60,000 Hz;

bats, 100,000 Hz; and dolphins 150,000 Hz, and American shad (*Alosa sapidissima*), a fish, can hear 180,000 Hz. Those frequencies above the human range are called **ultrasound**.

Amplitude, or the dimension of a wave from peak to trough, in sound is heard as volume and is illustrated in [link]. The sound waves of louder sounds have greater amplitude than those of softer sounds. For sound, volume is measured in decibels (dB). The softest sound that a human can hear is the zero point. Humans speak normally at 60 decibels.

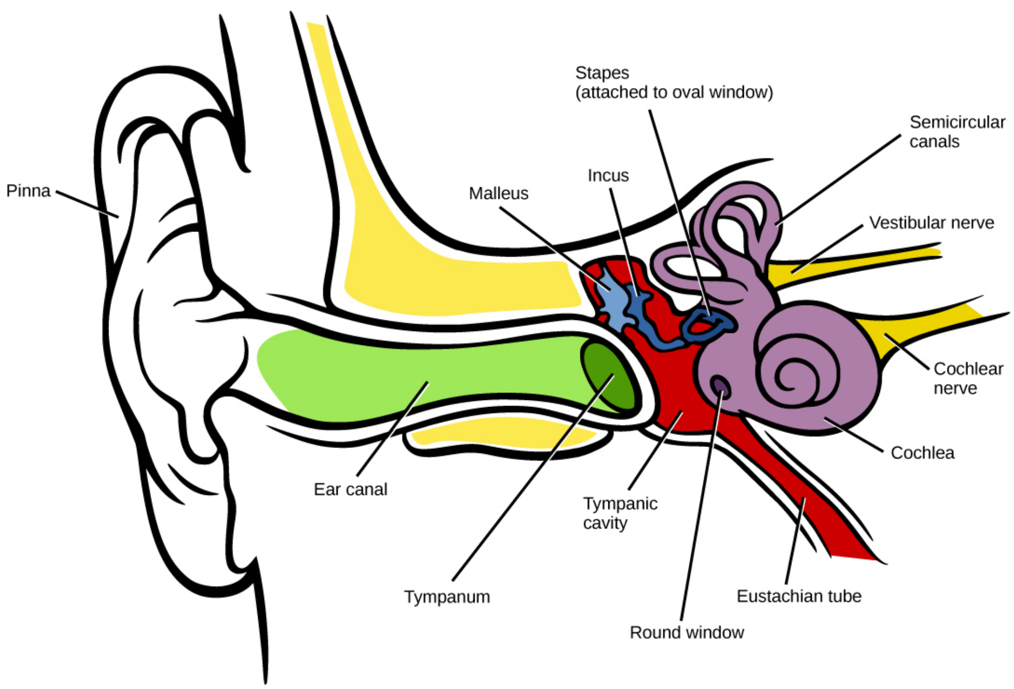


For sound waves, wavelength corresponds to pitch. Amplitude of the wave corresponds to volume. The sound wave shown with a dashed line is softer in volume than the sound wave shown with a solid line. (credit: NIH)

Reception of Sound

In mammals, sound waves are collected by the external, cartilaginous part of the ear called the **pinna**, then travel through the auditory canal and cause

vibration of the thin diaphragm called the **tympanum** or ear drum, the innermost part of the **outer ear** (illustrated in [link]). Interior to the tympanum is the **middle ear**. The middle ear holds three small bones called the **ossicles**, which transfer energy from the moving tympanum to the inner ear. The three ossicles are the **malleus** (also known as the hammer), the **incus** (the anvil), and **stapes** (the stirrup). The aptly named stapes looks very much like a stirrup. The three ossicles are unique to mammals, and each plays a role in hearing. The malleus attaches at three points to the interior surface of the tympanic membrane. The incus attaches the malleus to the stapes. In humans, the stapes is not long enough to reach the tympanum. If we did not have the malleus and the incus, then the vibrations of the tympanum would never reach the inner ear. These bones also function to collect force and amplify sounds. The ear ossicles are homologous to bones in a fish mouth: the bones that support gills in fish are thought to be adapted for use in the vertebrate ear over evolutionary time. Many animals (frogs, reptiles, and birds, for example) use the stapes of the middle ear to transmit vibrations to the middle ear.



Sound travels through the outer ear to the middle ear, which is bounded on its exterior by the tympanic membrane. The

middle ear contains three bones called ossicles that transfer the sound wave to the oval window, the exterior boundary of the inner ear. The organ of Corti, which is the organ of sound transduction, lies inside the cochlea. (credit: modification of work by Lars Chittka, Axel Brockmann)

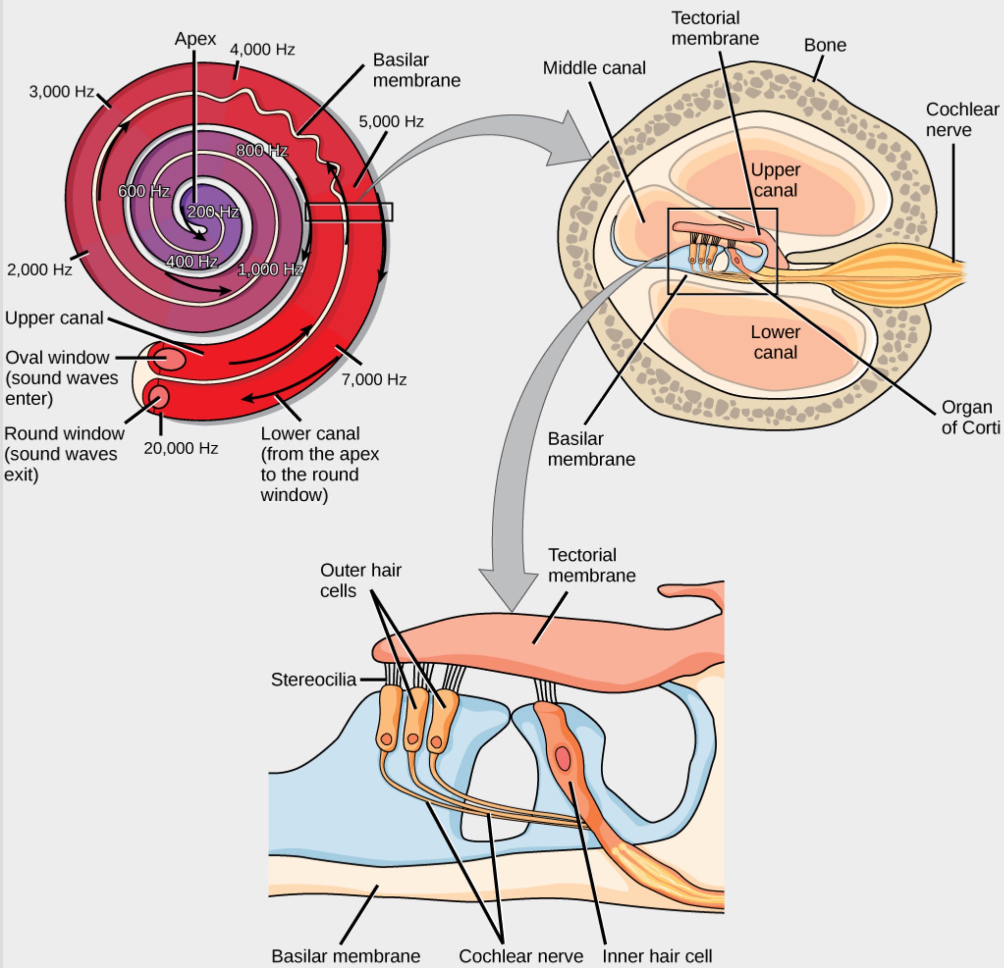
Transduction of Sound

Vibrating objects, such as vocal cords, create sound waves or pressure waves in the air. When these pressure waves reach the ear, the ear transduces this mechanical stimulus (pressure wave) into a nerve impulse (electrical signal) that the brain perceives as sound. The pressure waves strike the tympanum, causing it to vibrate. The mechanical energy from the moving tympanum transmits the vibrations to the three bones of the middle ear. The stapes transmits the vibrations to a thin diaphragm called the **oval window**, which is the outermost structure of the **inner ear**. The structures of the inner ear are found in the **labyrinth**, a bony, hollow structure that is the most interior portion of the ear. Here, the energy from the sound wave is transferred from the stapes through the flexible oval window and to the fluid of the cochlea. The vibrations of the oval window create pressure waves in the fluid (perilymph) inside the cochlea. The **cochlea** is a whorled structure, like the shell of a snail, and it contains receptors for transduction of the mechanical wave into an electrical signal (as illustrated in [link]). Inside the cochlea, the **basilar membrane** is a mechanical analyzer that runs the length of the cochlea, curling toward the cochlea’s center.

The mechanical properties of the basilar membrane change along its length, such that it is thicker, tauter, and narrower at the outside of the whorl (where the cochlea is largest), and thinner, floppier, and broader toward the apex, or center, of the whorl (where the cochlea is smallest). Different regions of the basilar membrane vibrate according to the frequency of the sound wave conducted through the fluid in the cochlea. For these reasons, the fluid-filled cochlea detects different wave frequencies (pitches) at different regions of the membrane. When the sound waves in the cochlear

fluid contact the basilar membrane, it flexes back and forth in a wave-like fashion. Above the basilar membrane is the **tectorial membrane**.

Note:
Art Connection



In the human ear, sound waves cause the stapes to press against the oval window. Vibrations travel up the fluid-filled interior of the cochlea. The basilar membrane that lines the cochlea gets continuously thinner toward the apex of the cochlea. Different thicknesses of membrane vibrate in response to different frequencies of sound. Sound waves then exit through the round window. In the cross section of the cochlea (top right figure), note that in

addition to the upper canal and lower canal, the cochlea also has a middle canal. The organ of Corti (bottom image) is the site of sound transduction. Movement of stereocilia on hair cells results in an action potential that travels along the auditory nerve.

Cochlear implants can restore hearing in people who have a nonfunctional cochlear. The implant consists of a microphone that picks up sound. A speech processor selects sounds in the range of human speech, and a transmitter converts these sounds to electrical impulses, which are then sent to the auditory nerve. Which of the following types of hearing loss would not be restored by a cochlear implant?

a. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
b. Hearing loss resulting from an abnormal auditory nerve.
c. Hearing loss resulting from fracture of the cochlea.
d. Hearing loss resulting from damage to bones of the middle ear.

The site of transduction is in the **organ of Corti** (spiral organ). It is composed of hair cells held in place above the basilar membrane like flowers projecting up from soil, with their exposed short, hair-like **stereocilia** contacting or embedded in the tectorial membrane above them. The inner hair cells are the primary auditory receptors and exist in a single row, numbering approximately 3,500. The stereocilia from inner hair cells extend into small dimples on the tectorial membrane’s lower surface. The outer hair cells are arranged in three or four rows. They number approximately 12,000, and they function to fine tune incoming sound waves. The longer stereocilia that project from the outer hair cells actually attach to the tectorial membrane. All of the stereocilia are mechanoreceptors, and when bent by vibrations they respond by opening a gated ion channel (refer to [link]). As a result, the hair cell membrane is depolarized, and a signal is transmitted to the chochlear nerve. Intensity

(volume) of sound is determined by how many hair cells at a particular location are stimulated.

The hair cells are arranged on the basilar membrane in an orderly way. The basilar membrane vibrates in different regions, according to the frequency of the sound waves impinging on it. Likewise, the hair cells that lay above it are most sensitive to a specific frequency of sound waves. Hair cells can respond to a small range of similar frequencies, but they require stimulation of greater intensity to fire at frequencies outside of their optimal range. The difference in response frequency between adjacent inner hair cells is about 0.2 percent. Compare that to adjacent piano strings, which are about six percent different. Place theory, which is the model for how biologists think pitch detection works in the human ear, states that high frequency sounds selectively vibrate the basilar membrane of the inner ear near the entrance port (the oval window). Lower frequencies travel farther along the membrane before causing appreciable excitation of the membrane. The basic pitch-determining mechanism is based on the location along the membrane where the hair cells are stimulated. The place theory is the first step toward an understanding of pitch perception. Considering the extreme pitch sensitivity of the human ear, it is thought that there must be some auditory “sharpening” mechanism to enhance the pitch resolution.

When sound waves produce fluid waves inside the cochlea, the basilar membrane flexes, bending the stereocilia that attach to the tectorial membrane. Their bending results in action potentials in the hair cells, and auditory information travels along the neural endings of the bipolar neurons of the hair cells (collectively, the auditory nerve) to the brain. When the hairs bend, they release an excitatory neurotransmitter at a synapse with a sensory neuron, which then conducts action potentials to the central nervous system. The cochlear branch of the vestibulocochlear cranial nerve sends information on hearing. The auditory system is very refined, and there is some modulation or “sharpening” built in. The brain can send signals back to the cochlea, resulting in a change of length in the outer hair cells, sharpening or dampening the hair cells’ response to certain frequencies.

Note:

Link to Learning

Watch an animation of sound entering the outer ear, moving through the ear structure, stimulating cochlear nerve impulses, and eventually sending signals to the temporal lobe.

Higher Processing

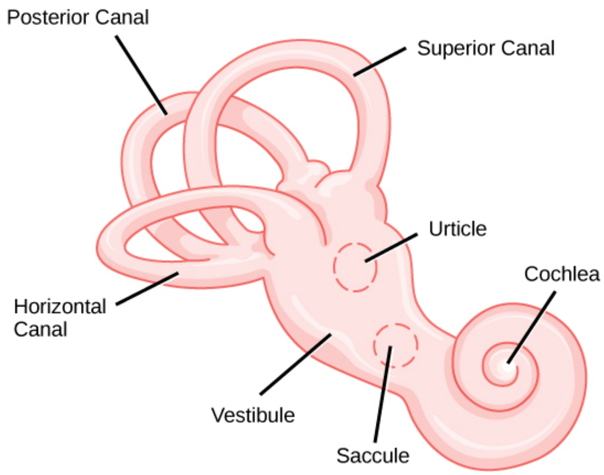
The inner hair cells are most important for conveying auditory information to the brain. About 90 percent of the afferent neurons carry information from inner hair cells, with each hair cell synapsing with 10 or so neurons. Outer hair cells connect to only 10 percent of the afferent neurons, and each afferent neuron innervates many hair cells. The afferent, bipolar neurons that convey auditory information travel from the cochlea to the medulla, through the pons and midbrain in the brainstem, finally reaching the primary auditory cortex in the temporal lobe.

Vestibular Information

The stimuli associated with the vestibular system are linear acceleration (gravity) and angular acceleration and deceleration. Gravity, acceleration, and deceleration are detected by evaluating the inertia on receptive cells in the vestibular system. Gravity is detected through head position. Angular acceleration and deceleration are expressed through turning or tilting of the head.

The vestibular system has some similarities with the auditory system. It utilizes hair cells just like the auditory system, but it excites them in different ways. There are five vestibular receptor organs in the inner ear: the

utricle, the saccule, and three semicircular canals. Together, they make up what's known as the vestibular labyrinth that is shown in [link]. The utricle and saccule respond to acceleration in a straight line, such as gravity. The roughly 30,000 hair cells in the utricle and 16,000 hair cells in the saccule lie below a gelatinous layer, with their stereocilia projecting into the gelatin. Embedded in this gelatin are calcium carbonate crystals—like tiny rocks. When the head is tilted, the crystals continue to be pulled straight down by gravity, but the new angle of the head causes the gelatin to shift, thereby bending the stereocilia. The bending of the stereocilia stimulates the neurons, and they signal to the brain that the head is tilted, allowing the maintenance of balance. It is the vestibular branch of the vestibulocochlear cranial nerve that deals with balance.



The structure of the vestibular labyrinth is shown. (credit: modification of work by NIH)

The fluid-filled **semicircular canals** are tubular loops set at oblique angles. They are arranged in three spatial planes. The base of each canal has a swelling that contains a cluster of hair cells. The hairs project into a gelatinous cap called the cupula and monitor angular acceleration and deceleration from rotation. They would be stimulated by driving your car

around a corner, turning your head, or falling forward. One canal lies horizontally, while the other two lie at about 45 degree angles to the horizontal axis, as illustrated in [link]. When the brain processes input from all three canals together, it can detect angular acceleration or deceleration in three dimensions. When the head turns, the fluid in the canals shifts, thereby bending stereocilia and sending signals to the brain. Upon cessation accelerating or decelerating—or just moving—the movement of the fluid within the canals slows or stops. For example, imagine holding a glass of water. When moving forward, water may splash backwards onto the hand, and when motion has stopped, water may splash forward onto the fingers. While in motion, the water settles in the glass and does not splash. Note that the canals are not sensitive to velocity itself, but to changes in velocity, so moving forward at 60mph with your eyes closed would not give the sensation of movement, but suddenly accelerating or braking would stimulate the receptors.

Higher Processing

Hair cells from the utricle, saccule, and semicircular canals also communicate through bipolar neurons to the cochlear nucleus in the medulla. Cochlear neurons send descending projections to the spinal cord and ascending projections to the pons, thalamus, and cerebellum. Connections to the cerebellum are important for coordinated movements. There are also projections to the temporal cortex, which account for feelings of dizziness; projections to autonomic nervous system areas in the brainstem, which account for motion sickness; and projections to the primary somatosensory cortex, which monitors subjective measurements of the external world and self-movement. People with lesions in the vestibular area of the somatosensory cortex see vertical objects in the world as being tilted. Finally, the vestibular signals project to certain optic muscles to coordinate eye and head movements.

Note:
Link to Learning

Click through this interactive tutorial to review the parts of the ear and how they function to process sound.

Section Summary

Audition is important for territory defense, predation, predator defense, and communal exchanges. The vestibular system, which is not auditory, detects linear acceleration and angular acceleration and deceleration. Both the auditory system and vestibular system use hair cells as their receptors.

Auditory stimuli are sound waves. The sound wave energy reaches the outer ear (pinna, canal, tympanum), and vibrations of the tympanum send the energy to the middle ear. The middle ear bones shift and the stapes transfers mechanical energy to the oval window of the fluid-filled inner ear cochlea. Once in the cochlea, the energy causes the basilar membrane to flex, thereby bending the stereocilia on receptor hair cells. This activates the receptors, which send their auditory neural signals to the brain.

The vestibular system has five parts that work together to provide the sense of direction, thus helping to maintain balance. The utricle and saccule measure head orientation: their calcium carbonate crystals shift when the head is tilted, thereby activating hair cells. The semicircular canals work similarly, such that when the head is turned, the fluid in the canals bends stereocilia on hair cells. The vestibular hair cells also send signals to the thalamus and to somatosensory cortex, but also to the cerebellum, the structure above the brainstem that plays a large role in timing and coordination of movement.

Art Connections

Exercise:

Problem:

[link] Cochlear implants can restore hearing in people who have a nonfunctional cochlear. The implant consists of a microphone that picks up sound. A speech processor selects sounds in the range of human speech, and a transmitter converts these sounds to electrical impulses, which are then sent to the auditory nerve. Which of the following types of hearing loss would not be restored by a cochlear implant?

a. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
b. Hearing loss resulting from an abnormal auditory nerve.
c. Hearing loss resulting from fracture of the cochlea.
d. Hearing loss resulting from damage to bones of the middle ear.

Solution:

[link] B

Review Questions

Exercise:

Problem:

In sound, pitch is measured in _____, and volume is measured in _____.

a. nanometers (nm); decibels (dB)
b. decibels (dB); nanometers (nm)
c. decibels (dB); hertz (Hz)
d. hertz (Hz); decibels (dB)

Solution:

D

Exercise:

Problem: Auditory hair cells are indirectly anchored to the ______.

a. basilar membrane
b. oval window
c. tectorial membrane
d. ossicles

Solution:

A

Exercise:

Problem:

Which of the following are found both in the auditory system and the vestibular system?

a. basilar membrane
b. hair cells
c. semicircular canals
d. ossicles

Solution:

B

Free Response

Exercise:

Problem:

How would a rise in altitude likely affect the speed of a sound transmitted through air? Why?

Solution:

The sound would slow down, because it is transmitted through the particles (gas) and there are fewer particles (lower density) at higher altitudes.

Exercise:

Problem:

How might being in a place with less gravity than Earth has (such as Earth's moon) affect vestibular sensation, and why?

Solution:

Because vestibular sensation relies on gravity's effects on tiny crystals in the inner ear, a situation of reduced gravity would likely impair vestibular sensation.

Glossary

audition
: sense of hearing

basilar membrane
: stiff structure in the cochlea that indirectly anchors auditory receptors

cochlea
: whorled structure that contains receptors for transduction of the mechanical wave into an electrical signal

incus
: (also, anvil) second of the three bones of the middle ear

inner ear
: innermost part of the ear; consists of the cochlea and the vestibular system

labyrinth
: bony, hollow structure that is the most internal part of the ear; contains the sites of transduction of auditory and vestibular information

malleus
: (also, hammer) first of the three bones of the middle ear

middle ear
: part of the hearing apparatus that functions to transfer energy from the tympanum to the oval window of the inner ear

organ of Corti
: in the basilar membrane, the site of the transduction of sound, a mechanical wave, to a neural signal

ossicle
: one of the three bones of the middle ear

outer ear
: part of the ear that consists of the pinna, ear canal, and tympanum and which conducts sound waves into the middle ear

oval window
: thin diaphragm between the middle and inner ears that receives sound waves from contact with the stapes bone of the middle ear

pinna
: cartilaginous outer ear

semicircular canal
: one of three half-circular, fluid-filled tubes in the vestibular labyrinth that monitors angular acceleration and deceleration

stapes

(also, stirrup) third of the three bones of the middle ear

stereocilia
: in the auditory system, hair-like projections from hair cells that help detect sound waves

tectorial membrane
: cochlear structure that lies above the hair cells and participates in the transduction of sound at the hair cells

tympanum
: (also, tympanic membrane or ear drum) thin diaphragm between the outer and middle ears

ultrasound
: sound frequencies above the human detectable ceiling of approximately 20,000 Hz

Vision

By the end of this section, you will be able to:

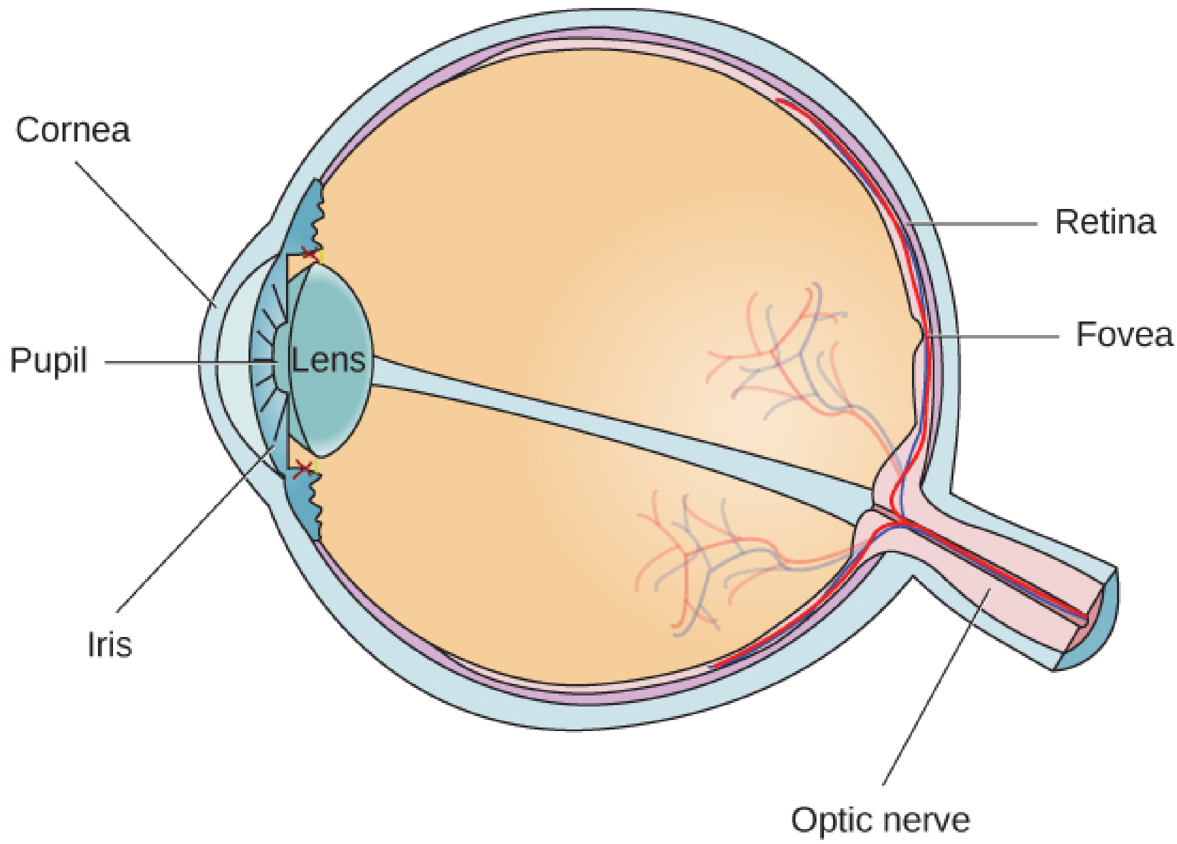
- Describe the basic anatomy of the visual system
- Discuss how rods and cones contribute to different aspects of vision
- Describe how monocular and binocular cues are used in the perception of depth

The visual system constructs a mental representation of the world around us ([link]). This contributes to our ability to successfully navigate through physical space and interact with important individuals and objects in our environments. This section will provide an overview of the basic anatomy and function of the visual system. In addition, we will explore our ability to perceive color and depth.

Our eyes take in sensory information that helps us understand the world around us. (credit "top left”: modification of work by "rajkumar1220"/Flickr"; credit “top right”: modification of work by Thomas Leuthard; credit “middle left”: modification of work by Demietrich Baker; credit “middle right”: modification of work by "kaybee07"/Flickr; credit “bottom left”: modification of work by "Isengardt"/Flickr; credit “bottom right”: modification of work by Willem Heerbaart)

ANATOMY OF THE VISUAL SYSTEM

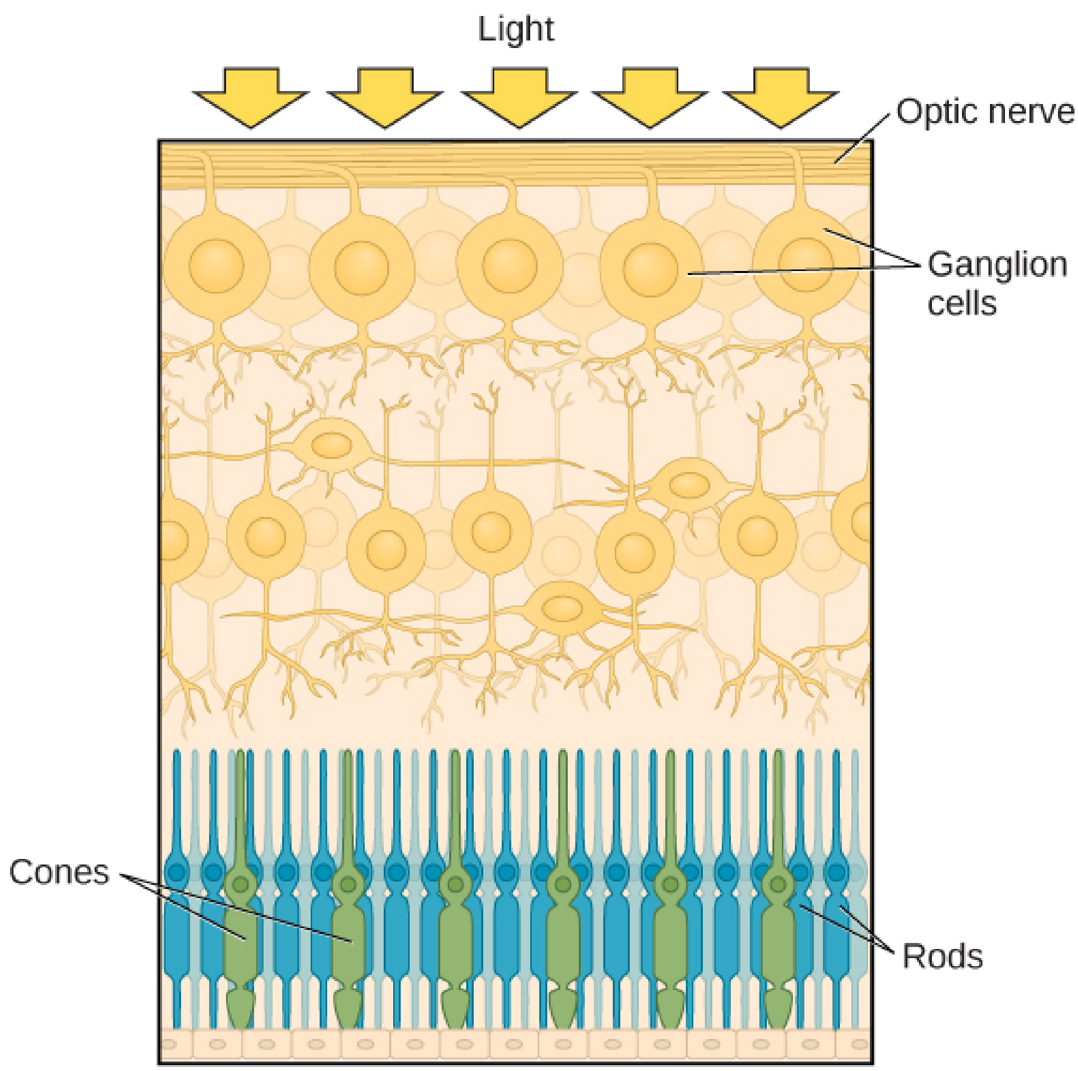
The eye is the major sensory organ involved in vision ([link]). Light waves are transmitted across the cornea and enter the eye through the pupil. The **cornea** is the transparent covering over the eye. It serves as a barrier between the inner eye and the outside world, and it is involved in focusing light waves that enter the eye. The **pupil** is the small opening in the eye through which light passes, and the size of the pupil can change as a function of light levels as well as emotional arousal. When light levels are low, the pupil will become dilated, or expanded, to allow more light to enter the eye. When light levels are high, the pupil will constrict, or become smaller, to reduce the amount of light that enters the eye. The pupil's size is controlled by muscles that are connected to the **iris**, which is the colored portion of the eye.



The anatomy of the eye is illustrated in this diagram.

After passing through the pupil, light crosses the **lens**, a curved, transparent structure that serves to provide additional focus. The lens is attached to muscles that can change its shape to aid in focusing light that is reflected from near or far objects. In a normal-sighted individual, the lens will focus images perfectly on a small indentation in the back of the eye known as the **fovea**, which is part of the **retina**, the light-sensitive lining of the eye. The fovea contains densely packed specialized photoreceptor cells ([link]). These **photoreceptor** cells, known as cones, are light-detecting cells. The **cones** are specialized types of photoreceptors that work best in bright light conditions. Cones are very sensitive to acute detail and provide tremendous spatial resolution. They also are directly involved in our ability to perceive color.

While cones are concentrated in the fovea, where images tend to be focused, rods, another type of photoreceptor, are located throughout the remainder of the retina. **Rods** are specialized photoreceptors that work well in low light conditions, and while they lack the spatial resolution and color function of the cones, they are involved in our vision in dimly lit environments as well as in our perception of movement on the periphery of our visual field.

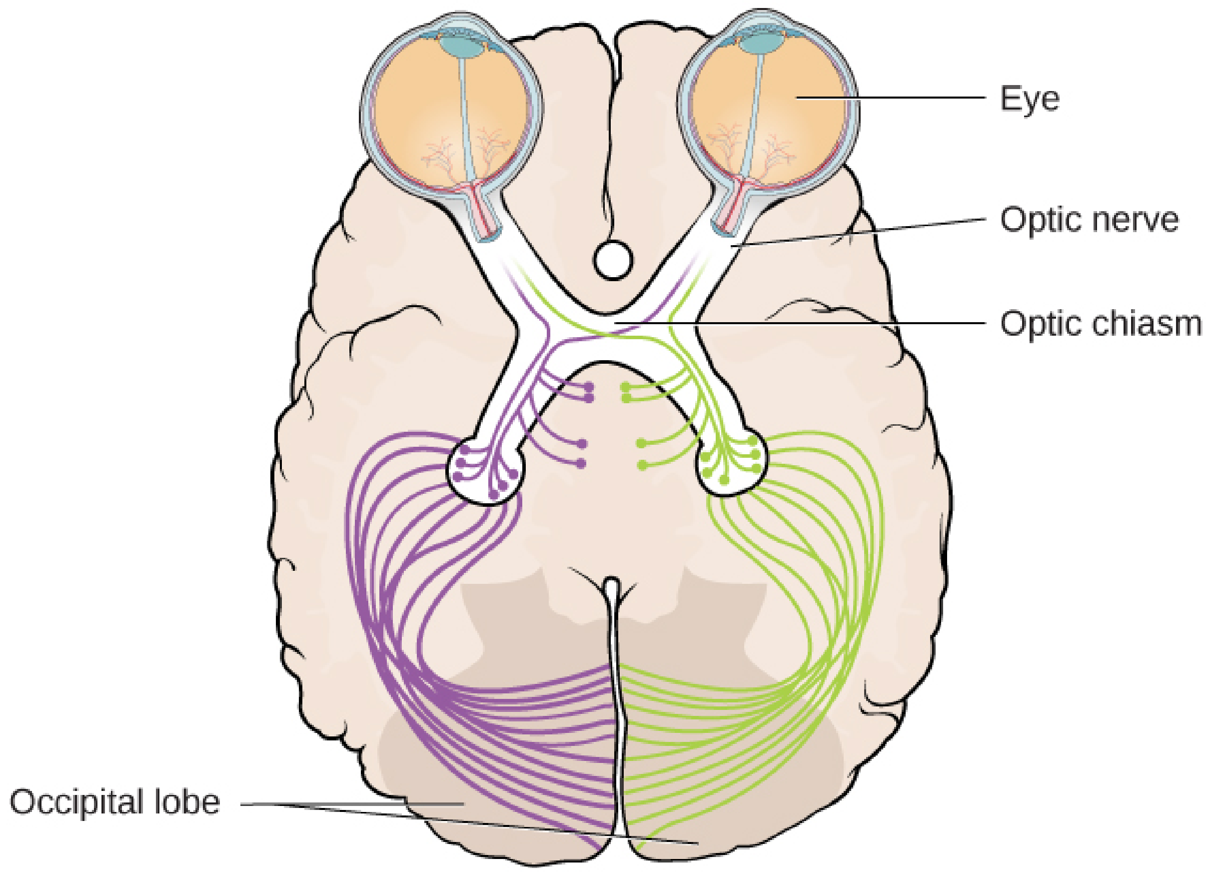


The two types of photoreceptors are shown in this image. Cones are colored green and rods are blue.

We have all experienced the different sensitivities of rods and cones when making the transition from a brightly lit environment to a dimly lit environment. Imagine going to see a blockbuster movie on a clear summer day. As you walk from the brightly lit lobby into the dark theater, you notice that you immediately have difficulty seeing much of anything. After a few minutes, you begin to adjust to the darkness and can see the interior of the theater. In the bright environment, your vision was dominated primarily by cone activity. As you move to the dark environment, rod activity dominates, but there is a delay in transitioning between the phases. If your rods do not transform light into nerve impulses as easily and efficiently as they should, you will have difficulty seeing in dim light, a condition known as night blindness.

Rods and cones are connected (via several interneurons) to retinal ganglion cells. Axons from the retinal ganglion cells converge and exit through the back of the eye to form the **optic nerve**. The optic nerve carries visual information from the retina to the brain. There is a point in the visual field called the **blind spot**: Even when light from a small object is focused on the blind spot, we do not see it. We are not consciously aware of our blind spots for two reasons: First, each eye gets a slightly different view of the visual field; therefore, the blind spots do not overlap. Second, our visual system fills in the blind spot so that although we cannot respond to visual information that occurs in that portion of the visual field, we are also not aware that information is missing.

The optic nerve from each eye merges just below the brain at a point called the **optic chiasm**. As [link] shows, the optic chiasm is an X-shaped structure that sits just below the cerebral cortex at the front of the brain. At the point of the optic chiasm, information from the right visual field (which comes from both eyes) is sent to the left side of the brain, and information from the left visual field is sent to the right side of the brain.



This illustration shows the optic chiasm at the front of the brain and the pathways to the occipital lobe at the back of the brain, where visual sensations are processed into meaningful perceptions.

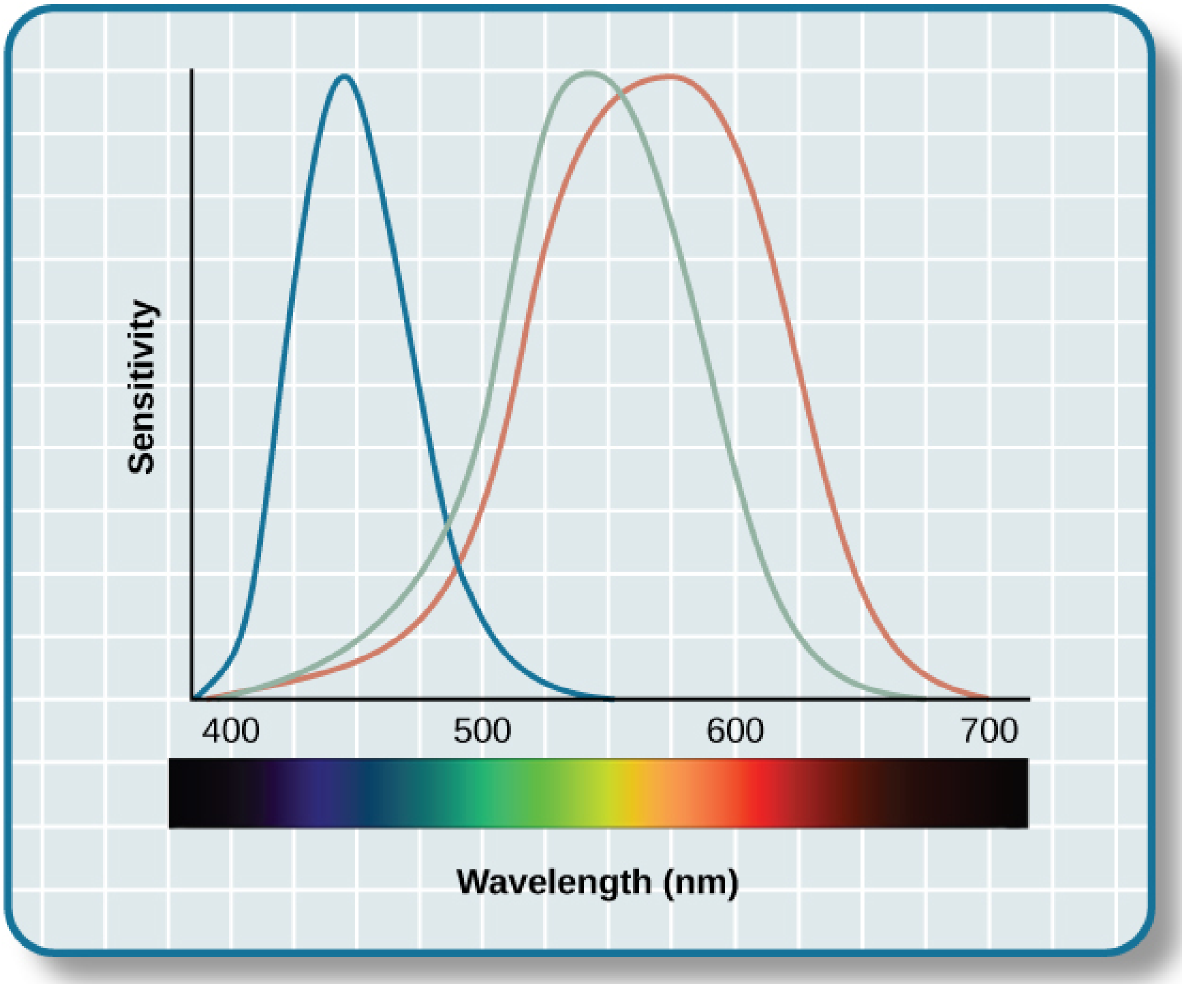
Once inside the brain, visual information is sent via a number of structures to the occipital lobe at the back of the brain for processing. Visual information might be processed in parallel pathways which can generally be described as the “what pathway” and the “where/how” pathway. The “what pathway” is involved in object recognition and identification, while the “where/how pathway” is involved with location in space and how one might interact with a particular visual stimulus (Milner & Goodale, 2008; Ungerleider & Haxby, 1994). For example, when you see a ball rolling down the street, the “what pathway” identifies what the object is, and the “where/how pathway” identifies its location or movement in space.

COLOR AND DEPTH PERCEPTION

We do not see the world in black and white; neither do we see it as two-dimensional (2-D) or flat (just height and width, no depth). Let’s look at how color vision works and how we perceive three dimensions (height, width, and depth).

Color Vision

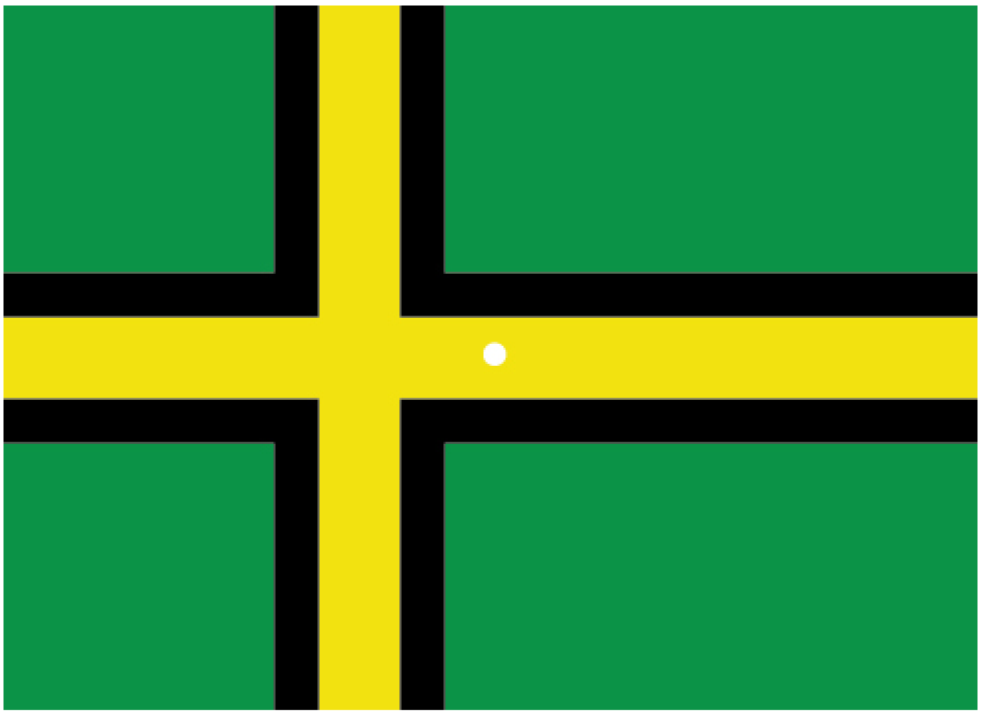
Normal-sighted individuals have three different types of cones that mediate color vision. Each of these cone types is maximally sensitive to a slightly different wavelength of light. According to the **trichromatic theory of color vision**, shown in [link], all colors in the spectrum can be produced by combining red, green, and blue. The three types of cones are each receptive to one of the colors.



This figure illustrates the different sensitivities for the three cone types found in a normal-sighted individual. (credit: modification of work by Vanessa Ezekowitz)

The trichromatic theory of color vision is not the only theory—another major theory of color vision is known as the **opponent-process theory**. According to this theory, color is coded in opponent pairs: black-white, yellow-blue, and green-red. The basic idea is that some cells of the visual system are excited by one of the opponent colors and inhibited by the other. So, a cell that was excited by wavelengths associated with green would be inhibited by wavelengths associated with red, and vice versa. One of the implications of opponent processing is that we do not experience greenish-

reds or yellowish-blues as colors. Another implication is that this leads to the experience of negative afterimages. An **afterimage** describes the continuation of a visual sensation after removal of the stimulus. For example, when you stare briefly at the sun and then look away from it, you may still perceive a spot of light although the stimulus (the sun) has been removed. When color is involved in the stimulus, the color pairings identified in the opponent-process theory lead to a negative afterimage. You can test this concept using the flag in [link].



Stare at the white dot for 30–60 seconds and then move your eyes to a blank piece of white paper. What do you see? This is known as a negative afterimage, and it provides empirical support for the opponent-process theory of color vision.

But these two theories—the trichromatic theory of color vision and the opponent-process theory—are not mutually exclusive. Research has shown

that they just apply to different levels of the nervous system. For visual processing on the retina, trichromatic theory applies: the cones are responsive to three different wavelengths that represent red, blue, and green. But once the signal moves past the retina on its way to the brain, the cells respond in a way consistent with opponent-process theory (Land, 1959; Kaiser, 1997).

Note:

Watch this video to learn about color vision in more detail.

Depth Perception

Our ability to perceive spatial relationships in three-dimensional (3-D) space is known as **depth perception**. With depth perception, we can describe things as being in front, behind, above, below, or to the side of other things.

Our world is three-dimensional, so it makes sense that our mental representation of the world has three-dimensional properties. We use a variety of cues in a visual scene to establish our sense of depth. Some of these are **binocular cues**, which means that they rely on the use of both eyes. One example of a binocular depth cue is **binocular disparity**, the slightly different view of the world that each of our eyes receives. To experience this slightly different view, do this simple exercise: extend your arm fully and extend one of your fingers and focus on that finger. Now, close your left eye without moving your head, then open your left eye and close your right eye without moving your head. You will notice that your

finger seems to shift as you alternate between the two eyes because of the slightly different view each eye has of your finger.

A 3-D movie works on the same principle: the special glasses you wear allow the two slightly different images projected onto the screen to be seen separately by your left and your right eye. As your brain processes these images, you have the illusion that the leaping animal or running person is coming right toward you.

Although we rely on binocular cues to experience depth in our 3-D world, we can also perceive depth in 2-D arrays. Think about all the paintings and photographs you have seen. Generally, you pick up on depth in these images even though the visual stimulus is 2-D. When we do this, we are relying on a number of **monocular cues**, or cues that require only one eye. If you think you can't see depth with one eye, note that you don't bump into things when using only one eye while walking—and, in fact, we have more monocular cues than binocular cues.

An example of a monocular cue would be what is known as linear perspective. **Linear perspective** refers to the fact that we perceive depth when we see two parallel lines that seem to converge in an image ([link]). Some other monocular depth cues are interposition, the partial overlap of objects, and the relative size and closeness of images to the horizon.

We perceive depth in a two-dimensional figure like this one through the use of monocular cues like linear perspective, like the parallel lines converging as the road narrows in the distance. (credit: Marc Dalmulder)

Note:
Stereoblindness
Bruce Bridgeman was born with an extreme case of lazy eye that resulted in him being stereoblind, or unable to respond to binocular cues of depth. He relied heavily on monocular depth cues, but he never had a true appreciation of the 3-D nature of the world around him. This all changed one night in 2012 while Bruce was seeing a movie with his wife.
The movie the couple was going to see was shot in 3-D, and even though he thought it was a waste of money, Bruce paid for the 3-D glasses when he purchased his ticket. As soon as the film began, Bruce put on the glasses and experienced something completely new. For the first time in his life he

appreciated the true depth of the world around him. Remarkably, his ability to perceive depth persisted outside of the movie theater.
There are cells in the nervous system that respond to binocular depth cues. Normally, these cells require activation during early development in order to persist, so experts familiar with Bruce's case (and others like his) assume that at some point in his development, Bruce must have experienced at least a fleeting moment of binocular vision. It was enough to ensure the survival of the cells in the visual system tuned to binocular cues. The mystery now is why it took Bruce nearly 70 years to have these cells activated (Peck, 2012).

Summary

Light waves cross the cornea and enter the eye at the pupil. The eye's lens focuses this light so that the image is focused on a region of the retina known as the fovea. The fovea contains cones that possess high levels of visual acuity and operate best in bright light conditions. Rods are located throughout the retina and operate best under dim light conditions. Visual information leaves the eye via the optic nerve. Information from each visual field is sent to the opposite side of the brain at the optic chiasm. Visual information then moves through a number of brain sites before reaching the occipital lobe, where it is processed.

Two theories explain color perception. The trichromatic theory asserts that three distinct cone groups are tuned to slightly different wavelengths of light, and it is the combination of activity across these cone types that results in our perception of all the colors we see. The opponent-process theory of color vision asserts that color is processed in opponent pairs and accounts for the interesting phenomenon of a negative afterimage. We perceive depth through a combination of monocular and binocular depth cues.

Review Questions

Exercise:

Problem:

The _________ is a small indentation of the retina that contains cones.

A. optic chiasm
B. optic nerve
C. fovea
D. iris

Solution:

C

Exercise:

Problem:_________ operate best under bright light conditions.

A. cones
B. rods
C. retinal ganglion cells
D. striate cortex

Solution:

A

Exercise:

Problem:_________ depth cues require the use of both eyes.

A. monocular
B. binocular
C. linear perspective
D. accommodating

Solution:

B

Exercise:

Problem:

If you were to stare at a green dot for a relatively long period of time and then shift your gaze to a blank white screen, you would see a _________ negative afterimage.

A. blue
B. yellow
C. black
D. red

Solution:

D

Critical Thinking Question

Exercise:

Problem:

Compare the two theories of color perception. Are they completely different?

Solution:

The trichromatic theory of color vision and the opponent-process theory are not mutually exclusive. Research has shown they apply to different levels of the nervous system. For visual processing on the retina, trichromatic theory applies: the cones are responsive to three different wavelengths that represent red, blue, and green. But once the

signal moves past the retina on its way to the brain, the cells respond in a way consistent with opponent-process theory.

Exercise:

Problem:

Color is not a physical property of our environment. What function (if any) do you think color vision serves?

Solution:

Color vision probably serves multiple adaptive purposes. One popular hypothesis suggests that seeing in color allowed our ancestors to differentiate ripened fruits and vegetables more easily.

Personal Application Question

Exercise:

Problem:

Take a look at a few of your photos or personal works of art. Can you find examples of linear perspective as a potential depth cue?

Glossary

afterimage
: continuation of a visual sensation after removal of the stimulus

binocular cue
: cue that relies on the use of both eyes

binocular disparity
: slightly different view of the world that each eye receives

blind spot

point where we cannot respond to visual information in that portion of the visual field

cone
: specialized photoreceptor that works best in bright light conditions and detects color

cornea
: transparent covering over the eye

depth perception
: ability to perceive depth

fovea
: small indentation in the retina that contains cones

iris
: colored portion of the eye

lens
: curved, transparent structure that provides additional focus for light entering the eye

linear perspective
: perceive depth in an image when two parallel lines seem to converge

monocular cue
: cue that requires only one eye

opponent-process theory of color perception
: color is coded in opponent pairs: black-white, yellow-blue, and red-green

optic chiasm
: X-shaped structure that sits just below the brain’s ventral surface; represents the merging of the optic nerves from the two eyes and the separation of information from the two sides of the visual field to the opposite side of the brain

optic nerve
carries visual information from the retina to the brain

photoreceptor
light-detecting cell

pupil
small opening in the eye through which light passes

retina
light-sensitive lining of the eye

rod
specialized photoreceptor that works well in low light conditions

trichromatic theory of color perception
color vision is mediated by the activity across the three groups of cones

Types of Hormones

By the end of this section, you will be able to:

- List the different types of hormones
- Explain their role in maintaining homeostasis

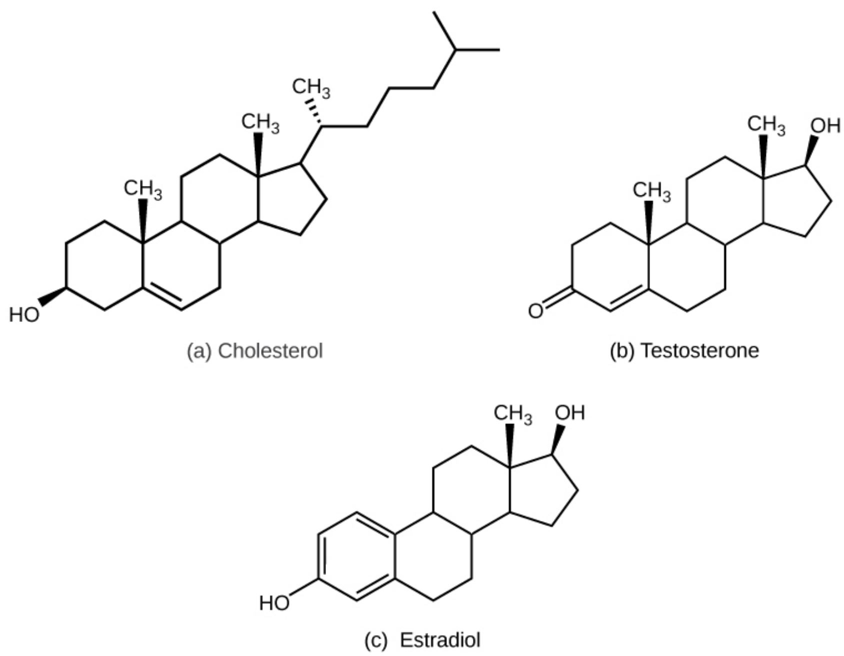
Maintaining homeostasis within the body requires the coordination of many different systems and organs. Communication between neighboring cells, and between cells and tissues in distant parts of the body, occurs through the release of chemicals called hormones. Hormones are released into body fluids (usually blood) that carry these chemicals to their target cells. At the target cells, which are cells that have a receptor for a signal or ligand from a signal cell, the hormones elicit a response. The cells, tissues, and organs that secrete hormones make up the endocrine system. Examples of glands of the endocrine system include the adrenal glands, which produce hormones such as epinephrine and norepinephrine that regulate responses to stress, and the thyroid gland, which produces thyroid hormones that regulate metabolic rates.

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure: lipid-derived, amino acid-derived, and peptide (peptide and proteins) hormones. One of the key distinguishing features of lipid-derived hormones is that they can diffuse across plasma membranes whereas the amino acid-derived and peptide hormones cannot.

Lipid-Derived Hormones (or Lipid-soluble Hormones)

Most **lipid hormones** are derived from cholesterol and thus are structurally similar to it, as illustrated in [link]. The primary class of lipid hormones in humans is the steroid hormones. Chemically, these hormones are usually ketones or alcohols; their chemical names will end in “-ol” for alcohols or “-one” for ketones. Examples of steroid hormones include estradiol, which is an **estrogen**, or female sex hormone, and testosterone, which is an androgen, or male sex hormone. These two hormones are released by the female and male reproductive organs, respectively. Other steroid hormones include aldosterone and cortisol, which are released by the adrenal glands

along with some other types of androgens. Steroid hormones are insoluble in water, and they are transported by transport proteins in blood. As a result, they remain in circulation longer than peptide hormones. For example, cortisol has a half-life of 60 to 90 minutes, while epinephrine, an amino acid derived-hormone, has a half-life of approximately one minute.

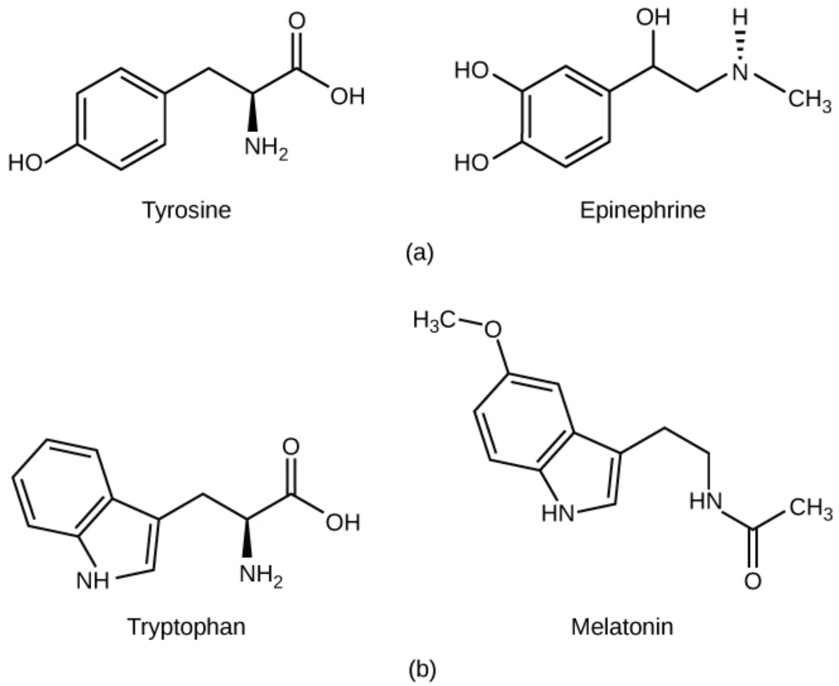


The structures shown here represent (a) cholesterol, plus the steroid hormones (b) testosterone and (c) estradiol.

Amino Acid-Derived Hormones

The **amino acid-derived hormones** are relatively small molecules that are derived from the amino acids tyrosine and tryptophan, shown in [link]. If a hormone is amino acid-derived, its chemical name will end in “-ine”. Examples of amino acid-derived hormones include epinephrine and norepinephrine, which are synthesized in the medulla of the adrenal glands,

and thyroxine, which is produced by the thyroid gland. The pineal gland in the brain makes and secretes melatonin which regulates sleep cycles.



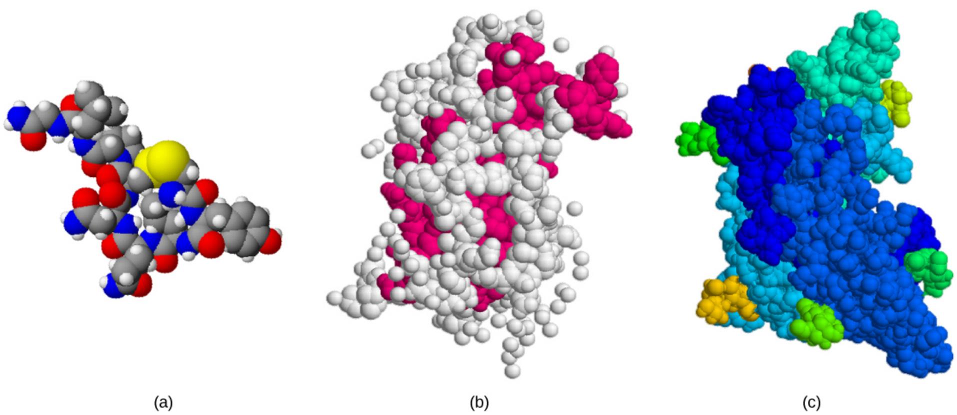
(a) The hormone epinephrine, which triggers the fight-or-flight response, is derived from the amino acid tyrosine. (b) The hormone melatonin, which regulates circadian rhythms, is derived from the amino acid tryptophan.

Peptide Hormones

The structure of **peptide hormones** is that of a polypeptide chain (chain of amino acids). The peptide hormones include molecules that are short polypeptide chains, such as antidiuretic hormone and oxytocin produced in the brain and released into the blood in the posterior pituitary gland. This class also includes small proteins, like growth hormones produced by the

pituitary, and large glycoproteins such as follicle-stimulating hormone produced by the pituitary. [link] illustrates these peptide hormones.

Secreted peptides like insulin are stored within vesicles in the cells that synthesize them. They are then released in response to stimuli such as high blood glucose levels in the case of insulin. Amino acid-derived and polypeptide hormones are water-soluble and insoluble in lipids. These hormones cannot pass through plasma membranes of cells; therefore, their receptors are found on the surface of the target cells.



The structures of peptide hormones (a) oxytocin, (b) growth hormone, and (c) follicle-stimulating hormone are shown. These peptide hormones are much larger than those derived from cholesterol or amino acids.

Note:
Career Connection
Endocrinologist
An endocrinologist is a medical doctor who specializes in treating disorders of the endocrine glands, hormone systems, and glucose and lipid metabolic pathways. An endocrine surgeon specializes in the surgical treatment of endocrine diseases and glands. Some of the diseases that are

managed by endocrinologists: disorders of the pancreas (diabetes mellitus), disorders of the pituitary (gigantism, acromegaly, and pituitary dwarfism), disorders of the thyroid gland (goiter and Graves' disease), and disorders of the adrenal glands (Cushing's disease and Addison's disease).

Endocrinologists are required to assess patients and diagnose endocrine disorders through extensive use of laboratory tests. Many endocrine diseases are diagnosed using tests that stimulate or suppress endocrine organ functioning. Blood samples are then drawn to determine the effect of stimulating or suppressing an endocrine organ on the production of hormones. For example, to diagnose diabetes mellitus, patients are required to fast for 12 to 24 hours. They are then given a sugary drink, which stimulates the pancreas to produce insulin to decrease blood glucose levels. A blood sample is taken one to two hours after the sugar drink is consumed. If the pancreas is functioning properly, the blood glucose level will be within a normal range. Another example is the A1C test, which can be performed during blood screening. The A1C test measures average blood glucose levels over the past two to three months by examining how well the blood glucose is being managed over a long time.

Once a disease has been diagnosed, endocrinologists can prescribe lifestyle changes and/or medications to treat the disease. Some cases of diabetes mellitus can be managed by exercise, weight loss, and a healthy diet; in other cases, medications may be required to enhance insulin release. If the disease cannot be controlled by these means, the endocrinologist may prescribe insulin injections.

In addition to clinical practice, endocrinologists may also be involved in primary research and development activities. For example, ongoing islet transplant research is investigating how healthy pancreas islet cells may be transplanted into diabetic patients. Successful islet transplants may allow patients to stop taking insulin injections.

Section Summary

There are three basic types of hormones: lipid-derived, amino acid-derived, and peptide. Lipid-derived hormones are structurally similar to cholesterol and include steroid hormones such as estradiol and testosterone. Amino

acid-derived hormones are relatively small molecules and include the adrenal hormones epinephrine and norepinephrine. Peptide hormones are polypeptide chains or proteins and include the pituitary hormones, antidiuretic hormone (vasopressin), and oxytocin.

Review Questions

Exercise:

Problem:

A newly discovered hormone contains four amino acids linked together. Under which chemical class would this hormone be classified?

a. lipid-derived hormone
b. amino acid-derived hormone
c. peptide hormone
d. glycoprotein

Solution:

C

Exercise:

Problem:

Which class of hormones can diffuse through plasma membranes?

a. lipid-derived hormones
b. amino acid-derived hormones
c. peptide hormones
d. glycoprotein hormones

Solution:

A

Free Response

Exercise:

Problem:

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure. What are these classes and what is one factor that distinguishes them?

Solution:

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure: lipid-derived, amino acid-derived, and peptide hormones. One of the key distinguishing features of the lipid-derived hormones is that they can diffuse across plasma membranes whereas the amino acid-derived and peptide hormones cannot.

Exercise:

Problem: Where is insulin stored, and why would it be released?

Solution:

Secreted peptides such as insulin are stored within vesicles in the cells that synthesize them. They are then released in response to stimuli such as high blood glucose levels in the case of insulin.

Glossary

amino acid-derived hormone
: hormone derived from amino acids

lipid-derived hormone

hormone derived mostly from cholesterol

peptide hormone
hormone composed of a polypeptide chain

Derived copy of How Hormones Work
By the end of this section, you will be able to:

- Explain how hormones work
- Discuss the role of different types of hormone receptors

Hormones mediate changes in target cells by binding to specific **hormone receptors**. In this way, even though hormones circulate throughout the body and come into contact with many different cell types, they only affect cells that possess the necessary receptors. Receptors for a specific hormone may be found on many different cells or may be limited to a small number of specialized cells. For example, thyroid hormones act on many different tissue types, stimulating metabolic activity throughout the body. Cells can have many receptors for the same hormone but often also possess receptors for different types of hormones. The number of receptors that respond to a hormone determines the cell’s sensitivity to that hormone, and the resulting cellular response. Additionally, the number of receptors that respond to a hormone can change over time, resulting in increased or decreased cell sensitivity. In **up-regulation**, the number of receptors increases in response to rising hormone levels, making the cell more sensitive to the hormone and allowing for more cellular activity. When the number of receptors decreases in response to rising hormone levels, called **down-regulation**, cellular activity is reduced.

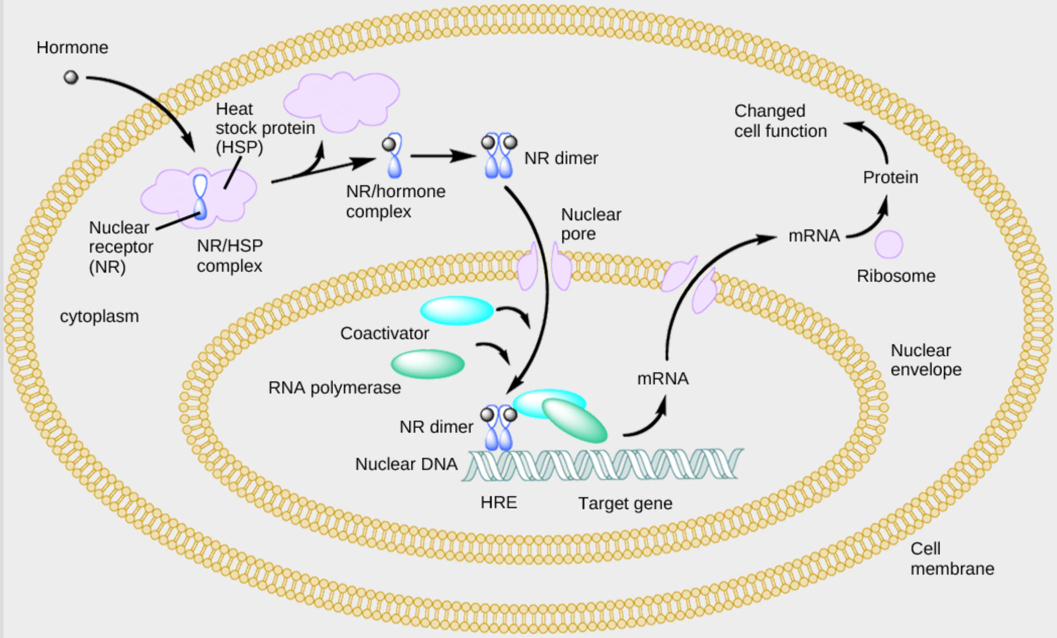
Receptor binding alters cellular activity and results in an increase or decrease in normal body processes. Depending on the location of the protein receptor on the target cell and the chemical structure of the hormone, hormones can mediate changes directly by binding to **intracellular hormone receptors** and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

Intracellular Hormone Receptors

Lipid-derived (soluble) hormones such as steroid hormones diffuse across the membranes of the endocrine cell. Once outside the cell, they bind to transport proteins that keep them soluble in the bloodstream. At the target

cell, the hormones are released from the carrier protein and diffuse across the lipid bilayer of the plasma membrane of cells. The steroid hormones pass through the plasma membrane of a target cell and adhere to intracellular receptors residing in the cytoplasm or in the nucleus. The cell signaling pathways induced by the steroid hormones regulate specific genes on the cell's DNA. The hormones and receptor complex act as transcription regulators by increasing or decreasing the synthesis of mRNA molecules of specific genes. This, in turn, determines the amount of corresponding protein that is synthesized by altering gene expression. This protein can be used either to change the structure of the cell or to produce enzymes that catalyze chemical reactions. In this way, the steroid hormone regulates specific cell processes as illustrated in [link].

Note:
Art Connection



An intracellular nuclear receptor (NR) is located in the cytoplasm bound to a heat shock protein (HSP). Upon hormone binding, the receptor dissociates from the heat shock protein and translocates to the nucleus. In the nucleus, the hormone-receptor complex binds to a DNA sequence called a hormone response element (HRE), which triggers

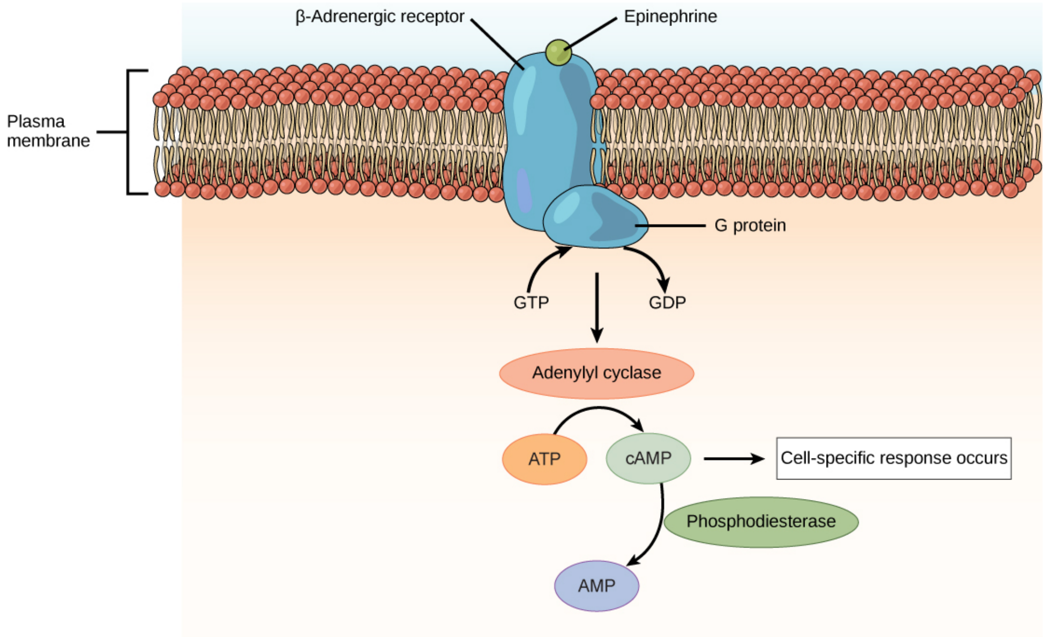
gene transcription and translation. The corresponding protein product can then mediate changes in cell function.

Heat shock proteins (HSP) are so named because they help refold misfolded proteins. In response to increased temperature (a “heat shock”), heat shock proteins are activated by release from the NR/HSP complex. At the same time, transcription of HSP genes is activated. Why do you think the cell responds to a heat shock by increasing the activity of proteins that help refold misfolded proteins?

Other lipid-soluble hormones that are not steroid hormones, such as vitamin D and thyroxine, have receptors located in the nucleus. The hormones diffuse across both the plasma membrane and the nuclear envelope, then bind to receptors in the nucleus. The hormone-receptor complex stimulates transcription of specific genes.

Plasma Membrane Hormone Receptors

Amino acid derived hormones and polypeptide hormones are not lipid-derived (lipid-soluble) and therefore cannot diffuse through the plasma membrane of cells. Lipid insoluble hormones bind to receptors on the outer surface of the plasma membrane, via **plasma membrane hormone receptors**. Unlike steroid hormones, lipid insoluble hormones do not directly affect the target cell because they cannot enter the cell and act directly on DNA. Binding of these hormones to a cell surface receptor results in activation of a signaling pathway; this triggers intracellular activity and carries out the specific effects associated with the hormone. In this way, nothing passes through the cell membrane; the hormone that binds at the surface remains at the surface of the cell while the intracellular product remains inside the cell. The hormone that initiates the signaling pathway is called a **first messenger**, which activates a second messenger in the cytoplasm, as illustrated in [link].



The amino acid-derived hormones epinephrine and norepinephrine bind to beta-adrenergic receptors on the plasma membrane of cells. Hormone binding to receptor activates a G-protein, which in turn activates adenylyl cyclase, converting ATP to cAMP. cAMP is a second messenger that mediates a cell-specific response. An enzyme called phosphodiesterase breaks down cAMP, terminating the signal.

The specific response of a cell to a lipid insoluble hormone depends on the type of receptors that are present on the cell membrane and the substrate molecules present in the cell cytoplasm. Cellular responses to hormone binding of a receptor include altering membrane permeability and metabolic pathways, stimulating synthesis of proteins and enzymes, and activating hormone release.

Section Summary

Hormones cause cellular changes by binding to receptors on target cells. The number of receptors on a target cell can increase or decrease in

response to hormone activity. Hormones can affect cells directly through intracellular hormone receptors or indirectly through plasma membrane hormone receptors.

Lipid-derived (soluble) hormones can enter the cell by diffusing across the plasma membrane and binding to DNA to regulate gene transcription and to change the cell’s activities by inducing production of proteins that affect, in general, the long-term structure and function of the cell. Lipid insoluble hormones bind to receptors on the plasma membrane surface and trigger a signaling pathway to change the cell’s activities by inducing production of various cell products that affect the cell in the short-term. The hormone is called a first messenger and the cellular component is called a second messenger. G-proteins activate the second messenger (cyclic AMP), triggering the cellular response. Response to hormone binding is amplified as the signaling pathway progresses. Cellular responses to hormones include the production of proteins and enzymes and altered membrane permeability.

Art Connections

Exercise:

Problem:

[link] Heat shock proteins (HSP) are so named because they help refold mis-folded proteins. In response to increased temperature (a “heat shock”), heat shock proteins are activated by release from the NR/HSP complex. At the same time, transcription of HSP genes is activated. Why do you think the cell responds to a heat shock by increasing the activity of proteins that help refold misfolded proteins?

Solution:

[link] Proteins unfold, or denature, at higher temperatures.

Review Questions

Exercise:

Problem:

A new antagonist molecule has been discovered that binds to and blocks plasma membrane receptors. What effect will this antagonist have on testosterone, a steroid hormone?

a. It will block testosterone from binding to its receptor.
b. It will block testosterone from activating cAMP signaling.
c. It will increase testosterone-mediated signaling.
d. It will not affect testosterone-mediated signaling.

Solution:

D

Free Response

Exercise:

Problem: Name two important functions of hormone receptors.

Solution:

The number of receptors that respond to a hormone can change, resulting in increased or decreased cell sensitivity. The number of receptors can increase in response to rising hormone levels, called up-regulation, making the cell more sensitive to the hormone and allowing for more cellular activity. The number of receptors can also decrease in response to rising hormone levels, called down-regulation, leading to reduced cellular activity.

Exercise:

Problem: How can hormones mediate changes?

Solution:

Depending on the location of the protein receptor on the target cell and the chemical structure of the hormone, hormones can mediate changes directly by binding to intracellular receptors and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

Glossary

adenylate cyclase
: an enzyme that catalyzes the conversion of ATP to cyclic AMP

down-regulation
: a decrease in the number of hormone receptors in response to increased hormone levels

first messenger
: the hormone that binds to a plasma membrane hormone receptor to trigger a signal transduction pathway

G-protein
: a membrane protein activated by the hormone first messenger to activate formation of cyclic AMP

hormone receptor
: the cellular protein that binds to a hormone

intracellular hormone receptor
: a hormone receptor in the cytoplasm or nucleus of a cell

phosphodiesterase (PDE)
: enzyme that deactivates cAMP, stopping hormone activity

plasma membrane hormone receptor
: a hormone receptor on the surface of the plasma membrane of a cell

up-regulation

an increase in the number of hormone receptors in response to increased hormone levels

Derived copy of Regulation of Body Processes
By the end of this section, you will be able to:

- Explain how hormones regulate the excretory system
- Discuss the role of hormones in the reproductive system
- Describe how hormones regulate metabolism
- Explain the role of hormones in different diseases

Hormones have a wide range of effects and modulate many different body processes. The key regulatory processes that will be examined here are those affecting the excretory system, the reproductive system, metabolism, blood calcium concentrations, growth, and the stress response.

Hormonal Regulation of the Excretory System

Maintaining a proper water balance in the body is important to avoid dehydration or over-hydration (hyponatremia). The water concentration of the body is monitored by **osmoreceptors** in the hypothalamus, which detect the concentration of electrolytes in the extracellular fluid. The concentration of electrolytes in the blood rises when there is water loss caused by excessive perspiration, inadequate water intake, or low blood volume due to blood loss. An increase in blood electrolyte levels results in a neuronal signal being sent from the osmoreceptors in hypothalamic nuclei. The pituitary gland has two components: anterior and posterior. The anterior pituitary is composed of glandular cells that secrete protein hormones. The posterior pituitary is an extension of the hypothalamus. It is composed largely of neurons that are continuous with the hypothalamus.

The hypothalamus produces a polypeptide hormone known as **antidiuretic hormone (ADH)**, which is transported to and released from the posterior pituitary gland. The principal action of ADH is to regulate the amount of water excreted by the kidneys. As ADH (which is also known as vasopressin) causes direct water reabsorption from the kidney tubules, salts and wastes are concentrated in what will eventually be excreted as urine. The hypothalamus controls the mechanisms of ADH secretion, either by regulating blood volume or the concentration of water in the blood. Dehydration or physiological stress can cause an increase of osmolarity

above 300 mOsm/L, which in turn, raises ADH secretion and water will be retained, causing an increase in blood pressure. ADH travels in the bloodstream to the kidneys. Once at the kidneys, ADH changes the kidneys to become more permeable to water by temporarily inserting water channels, aquaporins, into the kidney tubules. Water moves out of the kidney tubules through the aquaporins, reducing urine volume. The water is reabsorbed into the capillaries lowering blood osmolarity back toward normal. As blood osmolarity decreases, a negative feedback mechanism reduces osmoreceptor activity in the hypothalamus, and ADH secretion is reduced. ADH release can be reduced by certain substances, including alcohol, which can cause increased urine production and dehydration.

Chronic underproduction of ADH or a mutation in the ADH receptor results in **diabetes insipidus**. If the posterior pituitary does not release enough ADH, water cannot be retained by the kidneys and is lost as urine. This causes increased thirst, but water taken in is lost again and must be continually consumed. If the condition is not severe, dehydration may not occur, but severe cases can lead to electrolyte imbalances due to dehydration.

Hormonal Regulation of the Reproductive System

Regulation of the reproductive system is a process that requires the action of hormones from the pituitary gland, the adrenal cortex, and the gonads. During puberty in both males and females, the hypothalamus produces gonadotropin-releasing hormone (GnRH), which stimulates the production and release of **follicle-stimulating hormone (FSH)** and luteinizing hormone (LH) from the anterior pituitary gland. These hormones regulate the gonads (testes in males and ovaries in females) and therefore are called **gonadotropins**. In both males and females, FSH stimulates gamete production and LH stimulates production of hormones by the gonads. An increase in gonad hormone levels inhibits GnRH production through a negative feedback loop.

Regulation of the Male Reproductive System

In males, FSH stimulates the maturation of sperm cells. FSH production is inhibited by the hormone inhibin, which is released by the testes. LH stimulates production of the sex hormones (**androgens**) by the interstitial cells of the testes and therefore is also called interstitial cell-stimulating hormone.

The most widely known androgen in males is testosterone. Testosterone promotes the production of sperm and masculine characteristics. The adrenal cortex also produces small amounts of testosterone precursor, although the role of this additional hormone production is not fully understood.

Note:
Everyday Connection

The Dangers of Synthetic Hormones

Professional baseball player Jason Giambi publically admitted to, and apologized for, his use of anabolic steroids supplied by a trainer. (credit: Bryce Edwards)

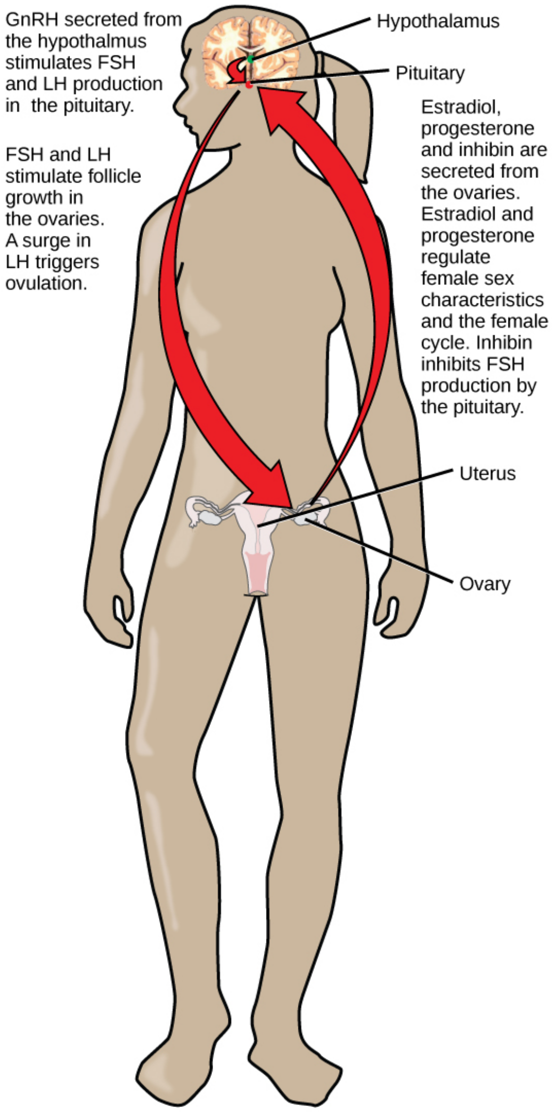
Some athletes attempt to boost their performance by using artificial hormones that enhance muscle performance. Anabolic steroids, a form of the male sex hormone testosterone, are one of the most widely known

performance-enhancing drugs. Steroids are used to help build muscle mass. Other hormones that are used to enhance athletic performance include erythropoietin, which triggers the production of red blood cells, and human growth hormone, which can help in building muscle mass. Most performance enhancing drugs are illegal for non-medical purposes. They are also banned by national and international governing bodies including the International Olympic Committee, the U.S. Olympic Committee, the National Collegiate Athletic Association, the Major League Baseball, and the National Football League.

The side effects of synthetic hormones are often significant and non-reversible, and in some cases, fatal. Androgens produce several complications such as liver dysfunctions and liver tumors, prostate gland enlargement, difficulty urinating, premature closure of epiphyseal cartilages, testicular atrophy, infertility, and immune system depression. The physiological strain caused by these substances is often greater than what the body can handle, leading to unpredictable and dangerous effects and linking their use to heart attacks, strokes, and impaired cardiac function.

Regulation of the Female Reproductive System

In females, FSH stimulates development of egg cells, called ova, which develop in structures called follicles. Follicle cells produce the hormone inhibin, which inhibits FSH production. LH also plays a role in the development of ova, induction of ovulation, and stimulation of estradiol and progesterone production by the ovaries, as illustrated in [link]. Estradiol and progesterone are steroid hormones that prepare the body for pregnancy. Estradiol produces secondary sex characteristics in females, while both estradiol and progesterone regulate the menstrual cycle.



Hormonal regulation of the female reproductive system involves hormones from the hypothalamus, pituitary, and ovaries.

In addition to producing FSH and LH, the anterior portion of the pituitary gland also produces the hormone **prolactin (PRL)** in females. Prolactin stimulates the production of milk by the mammary glands following childbirth. Prolactin levels are regulated by the hypothalamic hormones

prolactin-releasing hormone (PRH) and **prolactin-inhibiting hormone (PIH)**, which is now known to be dopamine. PRH stimulates the release of prolactin and PIH inhibits it.

The posterior pituitary releases the hormone **oxytocin**, which stimulates uterine contractions during childbirth. The uterine smooth muscles are not very sensitive to oxytocin until late in pregnancy when the number of oxytocin receptors in the uterus peaks. Stretching of tissues in the uterus and cervix stimulates oxytocin release during childbirth. Contractions increase in intensity as blood levels of oxytocin rise via a positive feedback mechanism until the birth is complete. Oxytocin also stimulates the contraction of myoepithelial cells around the milk-producing mammary glands. As these cells contract, milk is forced from the secretory alveoli into milk ducts and is ejected from the breasts in milk ejection ("let-down") reflex. Oxytocin release is stimulated by the suckling of an infant, which triggers the synthesis of oxytocin in the hypothalamus and its release into circulation at the posterior pituitary.

Hormonal Regulation of Metabolism

Blood glucose levels vary widely over the course of a day as periods of food consumption alternate with periods of fasting. Insulin and glucagon are the two hormones primarily responsible for maintaining homeostasis of blood glucose levels. Additional regulation is mediated by the thyroid hormones.

Regulation of Blood Glucose Levels by Insulin and Glucagon

Cells of the body require nutrients in order to function, and these nutrients are obtained through feeding. In order to manage nutrient intake, storing excess intake and utilizing reserves when necessary, the body uses hormones to moderate energy stores. **Insulin** is produced by the beta cells of the pancreas, which are stimulated to release insulin as blood glucose levels rise (for example, after a meal is consumed). Insulin lowers blood glucose levels by enhancing the rate of glucose uptake and utilization by

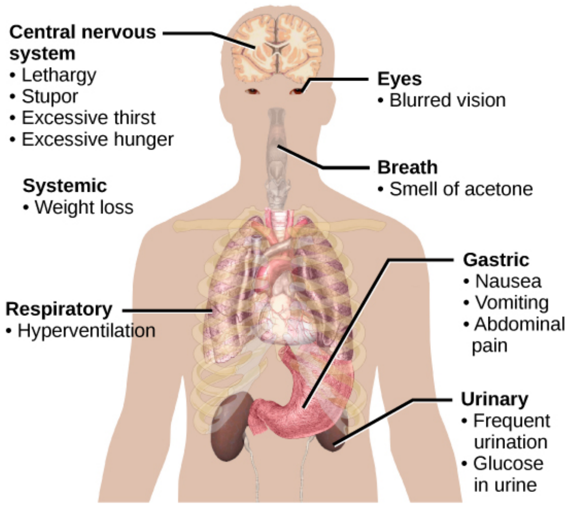
target cells, which use glucose for ATP production. It also stimulates the liver to convert glucose to glycogen, which is then stored by cells for later use. Insulin also increases glucose transport into certain cells, such as muscle cells and the liver. This results from an insulin-mediated increase in the number of glucose transporter proteins in cell membranes, which remove glucose from circulation by facilitated diffusion. As insulin binds to its target cell via insulin receptors and signal transduction, it triggers the cell to incorporate glucose transport proteins into its membrane. This allows glucose to enter the cell, where it can be used as an energy source. However, this does not occur in all cells: some cells, including those in the kidneys and brain, can access glucose without the use of insulin. Insulin also stimulates the conversion of glucose to fat in adipocytes and the synthesis of proteins. These actions mediated by insulin cause blood glucose concentrations to fall, called a hypoglycemic "low sugar" effect, which inhibits further insulin release from beta cells through a negative feedback loop.

Note:
Link to Learning

This animation describe the role of insulin and the pancreas in diabetes.

Impaired insulin function can lead to a condition called **diabetes mellitus**, the main symptoms of which are illustrated in [link]. This can be caused by low levels of insulin production by the beta cells of the pancreas, or by reduced sensitivity of tissue cells to insulin. This prevents glucose from being absorbed by cells, causing high levels of blood glucose, or **hyperglycemia** (high sugar). High blood glucose levels make it difficult for

the kidneys to recover all the glucose from nascent urine, resulting in glucose being lost in urine. High glucose levels also result in less water being reabsorbed by the kidneys, causing high amounts of urine to be produced; this may result in dehydration. Over time, high blood glucose levels can cause nerve damage to the eyes and peripheral body tissues, as well as damage to the kidneys and cardiovascular system. Oversecretion of insulin can cause **hypoglycemia**, low blood glucose levels. This causes insufficient glucose availability to cells, often leading to muscle weakness, and can sometimes cause unconsciousness or death if left untreated.

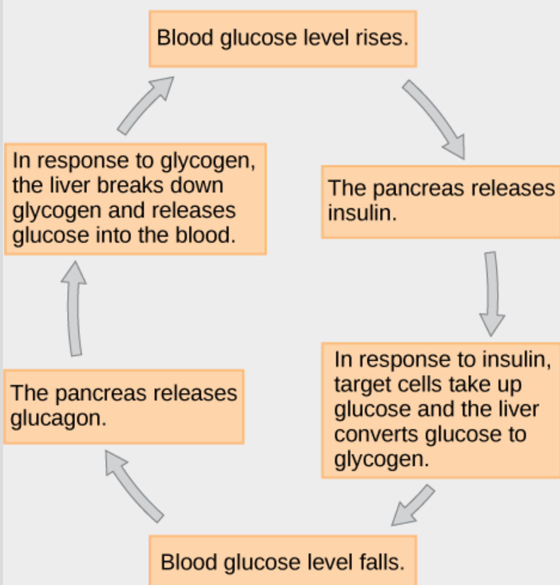


The main symptoms of diabetes are shown. (credit: modification of work by Mikael Häggström)

When blood glucose levels decline below normal levels, for example between meals or when glucose is utilized rapidly during exercise, the hormone **glucagon** is released from the alpha cells of the pancreas. Glucagon raises blood glucose levels, eliciting what is called a hyperglycemic effect, by stimulating the breakdown of glycogen to glucose in skeletal muscle cells and liver cells in a process called **glycogenolysis**.

Glucose can then be utilized as energy by muscle cells and released into circulation by the liver cells. Glucagon also stimulates absorption of amino acids from the blood by the liver, which then converts them to glucose. This process of glucose synthesis is called **gluconeogenesis**. Glucagon also stimulates adipose cells to release fatty acids into the blood. These actions mediated by glucagon result in an increase in blood glucose levels to normal homeostatic levels. Rising blood glucose levels inhibit further glucagon release by the pancreas via a negative feedback mechanism. In this way, insulin and glucagon work together to maintain homeostatic glucose levels, as shown in [link].

Note:
Art Connection



Insulin and glucagon regulate blood glucose levels.

Pancreatic tumors may cause excess secretion of glucagon. Type I diabetes results from the failure of the pancreas to produce insulin. Which of the following statement about these two conditions is true?

a. A pancreatic tumor and type I diabetes will have the opposite effects on blood sugar levels.
b. A pancreatic tumor and type I diabetes will both cause hyperglycemia.
c. A pancreatic tumor and type I diabetes will both cause hypoglycemia.
d. Both pancreatic tumors and type I diabetes result in the inability of cells to take up glucose.

Hormonal Regulation of Stress

When a threat or danger is perceived, the body responds by releasing hormones that will ready it for the "fight-or-flight" response. The effects of this response are familiar to anyone who has been in a stressful situation: increased heart rate, dry mouth, and hair standing up.

Note:
Evolution Connection
Fight-or-Flight Response
Interactions of the endocrine hormones have evolved to ensure the body's internal environment remains stable. Stressors are stimuli that disrupt homeostasis. The sympathetic division of the vertebrate autonomic nervous system has evolved the fight-or-flight response to counter stress-induced disruptions of homeostasis. In the initial alarm phase, the sympathetic nervous system stimulates an increase in energy levels through increased blood glucose levels. This prepares the body for physical activity that may be required to respond to stress: to either fight for survival or to flee from danger.
However, some stresses, such as illness or injury, can last for a long time. Glycogen reserves, which provide energy in the short-term response to stress, are exhausted after several hours and cannot meet long-term energy needs. If glycogen reserves were the only energy source available, neural functioning could not be maintained once the reserves became depleted due to the nervous system's high requirement for glucose. In this situation, the body has evolved a response to counter long-term stress through the

actions of the glucocorticoids, which ensure that long-term energy requirements can be met. The glucocorticoids mobilize lipid and protein reserves, stimulate gluconeogenesis, conserve glucose for use by neural tissue, and stimulate the conservation of salts and water. The mechanisms to maintain homeostasis that are described here are those observed in the human body. However, the fight-or-flight response exists in some form in all vertebrates.

The sympathetic nervous system regulates the stress response via the hypothalamus. Stressful stimuli cause the hypothalamus to signal the adrenal medulla (which mediates short-term stress responses) via nerve impulses, and the adrenal cortex, which mediates long-term stress responses, via the hormone **adrenocorticotropic hormone (ACTH)**, which is produced by the anterior pituitary.

Short-term Stress Response

When presented with a stressful situation, the body responds by calling for the release of hormones that provide a burst of energy. The hormones **epinephrine** (also known as adrenaline) and **norepinephrine** (also known as noradrenaline) are released by the adrenal medulla. How do these hormones provide a burst of energy? Epinephrine and norepinephrine increase blood glucose levels by stimulating the liver and skeletal muscles to break down glycogen and by stimulating glucose release by liver cells. Additionally, these hormones increase oxygen availability to cells by increasing the heart rate and dilating the bronchioles. The hormones also prioritize body function by increasing blood supply to essential organs such as the heart, brain, and skeletal muscles, while restricting blood flow to organs not in immediate need, such as the skin, digestive system, and kidneys. Epinephrine and norepinephrine are collectively called catecholamines.

Note:
Link to Learning

Watch this Discovery Channel animation describing the flight-or-flight response.

Long-term Stress Response

Long-term stress response differs from short-term stress response. The body cannot sustain the bursts of energy mediated by epinephrine and norepinephrine for long times. Instead, other hormones come into play. In a long-term stress response, the hypothalamus triggers the release of ACTH from the anterior pituitary gland. The adrenal cortex is stimulated by ACTH to release steroid hormones called **corticosteroids**. Corticosteroids turn on transcription of certain genes in the nuclei of target cells. They change enzyme concentrations in the cytoplasm and affect cellular metabolism. There are two main corticosteroids: glucocorticoids such as **cortisol**, and mineralocorticoids such as aldosterone. These hormones target the breakdown of fat into fatty acids in the adipose tissue. The fatty acids are released into the bloodstream for other tissues to use for ATP production. The **glucocorticoids** primarily affect glucose metabolism by stimulating glucose synthesis. Glucocorticoids also have anti-inflammatory properties through inhibition of the immune system. For example, cortisone is used as an anti-inflammatory medication; however, it cannot be used long term as it increases susceptibility to disease due to its immune-suppressing effects.

Mineralocorticoids function to regulate ion and water balance of the body. The hormone aldosterone stimulates the reabsorption of water and sodium ions in the kidney, which results in increased blood pressure and volume.

Hypersecretion of glucocorticoids can cause a condition known as **Cushing’s disease**, characterized by a shifting of fat storage areas of the body. This can cause the accumulation of adipose tissue in the face and neck, and excessive glucose in the blood. Hyposecretion of the corticosteroids can cause **Addison’s disease**, which may result in bronzing of the skin, hypoglycemia, and low electrolyte levels in the blood.

Section Summary

Water levels in the body are controlled by antidiuretic hormone (ADH), which is produced in the hypothalamus and triggers the reabsorption of water by the kidneys. Underproduction of ADH can cause diabetes insipidus. Aldosterone, a hormone produced by the adrenal cortex of the kidneys, enhances Na^+ reabsorption from the extracellular fluids and subsequent water reabsorption by diffusion. The renin-angiotensin-aldosterone system is one way that aldosterone release is controlled.

The reproductive system is controlled by the gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which are produced by the pituitary gland. Gonadotropin release is controlled by the hypothalamic hormone gonadotropin-releasing hormone (GnRH). FSH stimulates the maturation of sperm cells in males and is inhibited by the hormone inhibin, while LH stimulates the production of the androgen testosterone. FSH stimulates egg maturation in females, while LH stimulates the production of estrogens and progesterone. **Estrogens** are a group of steroid hormones produced by the ovaries that trigger the development of secondary sex characteristics in females as well as control the maturation of the ova. In females, the pituitary also produces prolactin, which stimulates milk production after childbirth, and oxytocin, which stimulates uterine contraction during childbirth and milk let-down during suckling.

Insulin is produced by the pancreas in response to rising blood glucose levels and allows cells to utilize blood glucose and store excess glucose for later use. Diabetes mellitus is caused by reduced insulin activity and causes high blood glucose levels, or hyperglycemia. Glucagon is released by the pancreas in response to low blood glucose levels and stimulates the

breakdown of glycogen into glucose, which can be used by the body. The body's basal metabolic rate is controlled by the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3). The anterior pituitary produces thyroid stimulating hormone (TSH), which controls the release of T_3 and T_4 from the thyroid gland. Iodine is necessary in the production of thyroid hormone, and the lack of iodine can lead to a condition called goiter.

Art Connections

Exercise:

Problem:

[link] Pancreatic tumors may cause excess secretion of glucagon. Type I diabetes results from the failure of the pancreas to produce insulin. Which of the following statement about these two conditions is true?

a. A pancreatic tumor and type I diabetes will have the opposite effects on blood sugar levels.
b. A pancreatic tumor and type I diabetes will both cause hyperglycemia.
c. A pancreatic tumor and type I diabetes will both cause hypoglycemia.
d. Both pancreatic tumors and type I diabetes result in the inability of cells to take up glucose.

Solution:

[link] B

Review Questions

Exercise:

Problem:

FSH and LH release from the anterior pituitary is stimulated by ________.

a. TSH
b. GnRH
c. T_3
d. PTH

Solution:

B

Exercise:

Problem: What hormone is produced by beta cells of the pancreas?

a. T_3
b. glucagon
c. insulin
d. T_4

Solution:

C

Free Response

Exercise:

Problem:

Name and describe a function of one hormone produced by the anterior pituitary and one hormone produced by the posterior pituitary.

Solution:

In addition to producing FSH and LH, the anterior pituitary also produces the hormone prolactin (PRL) in females. Prolactin stimulates the production of milk by the mammary glands following childbirth. Prolactin levels are regulated by the hypothalamic hormones prolactin-releasing hormone (PRH) and prolactin-inhibiting hormone (PIH) which is now known to be dopamine. PRH stimulates the release of prolactin and PIH inhibits it. The posterior pituitary releases the hormone oxytocin, which stimulates contractions during childbirth. The uterine smooth muscles are not very sensitive to oxytocin until late in pregnancy when the number of oxytocin receptors in the uterus peaks. Stretching of tissues in the uterus and vagina stimulates oxytocin release in childbirth. Contractions increase in intensity as blood levels of oxytocin rise until the birth is complete.

Glossary

acromegaly
: condition caused by overproduction of GH in adults

Addison’s disease
: disorder caused by the hyposecretion of corticosteroids

adrenocorticotropic hormone (ACTH)
: hormone released by the anterior pituitary, which stimulates the adrenal cortex to release corticosteroids during the long-term stress response

aldosterone
: steroid hormone produced by the adrenal cortex that stimulates the reabsorption of Na^+ from extracellular fluids and secretion of K^+.

androgen
: male sex hormone such as testosterone

antidiuretic hormone (ADH)

hormone produced by the hypothalamus and released by the posterior pituitary that increases water reabsorption by the kidneys

calcitonin
: hormone produced by the parafollicular cells of the thyroid gland that functions to lower blood Ca^{2+} levels and promote bone growth

corticosteroid
: hormone released by the adrenal cortex in response to long-term stress

cortisol
: glucocorticoid produced in response to stress

Cushing's disease
: disorder caused by the hypersecretion of glucocorticoids

diabetes insipidus
: disorder caused by underproduction of ADH

diabetes mellitus
: disorder caused by low levels of insulin activity

diabetogenic effect
: effect of GH that causes blood glucose levels to rise similar to diabetes mellitus

epinephrine
: hormone released by the adrenal medulla in response to a short term stress

estrogens
: - a group of steroid hormones, including estradiol and several others, that are produced by the ovaries and elicit secondary sex characteristics in females as well as control the maturation of the ova

follicle-stimulating hormone (FSH)
: hormone produced by the anterior pituitary that stimulates gamete production

gigantism
condition caused by overproduction of GH in children

glucagon
hormone produced by the alpha cells of the pancreas in response to low blood sugar; functions to raise blood sugar levels

glucocorticoid
corticosteroid that affects glucose metabolism

gluconeogenesis
synthesis of glucose from amino acids

glucose-sparing effect
effect of GH that causes tissues to use fatty acids instead of glucose as an energy source

glycogenolysis
breakdown of glycogen into glucose

goiter
enlargement of the thyroid gland caused by insufficient dietary iodine levels

gonadotropin
hormone that regulates the gonads, including FSH and LH

growth hormone (GH)
hormone produced by the anterior pituitary that promotes protein synthesis and body growth

growth hormone-inhibiting hormone (GHIH)
hormone produced by the hypothalamus that inhibits growth hormone production, also called somatostatin

growth hormone-releasing hormone (GHRH)
hormone released by the hypothalamus that triggers the release of GH

hyperglycemia

high blood sugar level

hyperthyroidism
overactivity of the thyroid gland

hypoglycemia
low blood sugar level

hypothyroidism
underactivity of the thyroid gland

insulin
hormone produced by the beta cells of the pancreas in response to high blood glucose levels; functions to lower blood glucose levels

insulin-like growth factor (IGF)
growth-promoting protein produced by the liver

mineralocorticoid
corticosteroid that affects ion and water balance

norepinephrine
hormone released by the adrenal medulla in response to a short-term stress hormone production by the gonads

osmoreceptor
receptor in the hypothalamus that monitors the concentration of electrolytes in the blood

oxytocin
hormone released by the posterior pituitary to stimulate uterine contractions during childbirth and milk let-down in the mammary glands

parathyroid hormone (PTH)
hormone produced by the parathyroid glands in response to low blood Ca^{2+} levels; functions to raise blood Ca^{2+} levels

pituitary dwarfism

condition caused by underproduction of GH in children

prolactin (PRL)
hormone produced by the anterior pituitary that stimulates milk production

prolactin-inhibiting hormone
hormone produced by the hypothalamus that inhibits the release of prolactin

prolactin-releasing hormone
hormone produced by the hypothalamus that stimulates the release of prolactin

renin
enzyme produced by the juxtaglomerular apparatus of the kidneys that reacts with angiotensinogen to cause the release of aldosterone

thyroglobulin
glycoprotein found in the thyroid that is converted into thyroid hormone

thyroid-stimulating hormone (TSH)
hormone produced by the anterior pituitary that controls the release of T_3 and T_4 from the thyroid gland

thyroxine (tetraiodothyronine, T_4)
thyroid hormone that controls the basal metabolic rate

triiodothyronine (T_3)
thyroid hormone that controls the basal metabolic rate

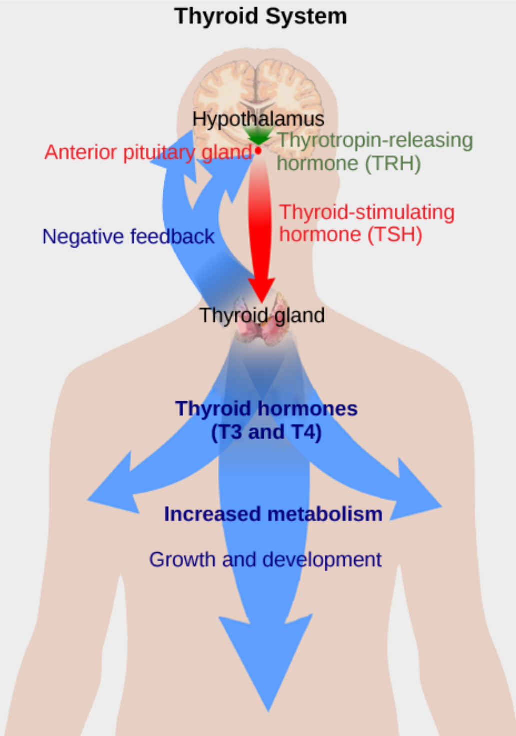
Regulation of Hormone Production

By the end of this section, you will be able to:

- Explain how hormone production is regulated
- Discuss the different stimuli that control hormone levels in the body

Hormone production and release are primarily controlled by negative feedback. In negative feedback systems, a stimulus elicits the release of a substance; once the substance reaches a certain level, it sends a signal that stops further release of the substance. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then give feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland, as illustrated in [link]. There are three mechanisms by which endocrine glands are stimulated to synthesize and release hormones: humoral stimuli, hormonal stimuli, and neural stimuli.

Note:
Art Connection



The anterior pituitary stimulates the thyroid gland to release thyroid hormones T_3 and T_4. Increasing levels of these hormones in the blood results in feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland. (credit: modification of work by Mikael Häggström)

Hyperthyroidism is a condition in which the thyroid gland is overactive. Hypothyroidism is a condition in which the thyroid gland is underactive. Which of the conditions are the following two patients most likely to have? Patient A has symptoms including weight gain, cold sensitivity, low heart rate and fatigue.

Patient B has symptoms including weight loss, profuse sweating, increased heart rate and difficulty sleeping.

Humoral Stimuli

The term “humoral” is derived from the term “humor,” which refers to bodily fluids such as blood. A **humoral stimulus** refers to the control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood. For example, a rise in blood glucose levels triggers the pancreatic release of insulin. Insulin causes blood glucose levels to drop, which signals the pancreas to stop producing insulin in a negative feedback loop.

Hormonal Stimuli

Hormonal stimuli refers to the release of a hormone in response to another hormone. A number of endocrine glands release hormones when stimulated by hormones released by other endocrine glands. For example, the hypothalamus produces hormones that stimulate the anterior portion of the pituitary gland. The anterior pituitary in turn releases hormones that regulate hormone production by other endocrine glands. The anterior pituitary releases the thyroid-stimulating hormone, which then stimulates the thyroid gland to produce the hormones T_3 and T_4. As blood concentrations of T_3 and T_4 rise, they inhibit both the pituitary and the hypothalamus in a negative feedback loop.

Neural Stimuli

In some cases, the nervous system directly stimulates endocrine glands to release hormones, which is referred to as **neural stimuli**. Recall that in a short-term stress response, the hormones epinephrine and norepinephrine are important for providing the bursts of energy required for the body to respond. Here, neuronal signaling from the sympathetic nervous system directly stimulates the adrenal medulla to release the hormones epinephrine and norepinephrine in response to stress.

Section Summary

Hormone levels are primarily controlled through negative feedback, in which rising levels of a hormone inhibit its further release. The three mechanisms of hormonal release are humoral stimuli, hormonal stimuli, and neural stimuli. Humoral stimuli refers to the control of hormonal release in response to changes in extracellular fluid levels or ion levels. Hormonal stimuli refers to the release of hormones in response to hormones released by other endocrine glands. Neural stimuli refers to the release of hormones in response to neural stimulation.

Art Connections

Exercise:

Problem:

[link] Hyperthyroidism is a condition in which the thyroid gland is overactive. Hypothyroidism is a condition in which the thyroid gland is underactive. Which of the conditions are the following two patients most likely to have?

Patient A has symptoms including weight gain, cold sensitivity, low heart rate and fatigue.

Patient B has symptoms including weight loss, profuse sweating, increased heart rate and difficulty sleeping.

Solution:

[link] Patient A has symptoms associated with decreased metabolism, and may be suffering from hypothyroidism. Patient B has symptoms associated with increased metabolism, and may be suffering from hyperthyroidism.

Review Questions

Exercise:

Problem:

A rise in blood glucose levels triggers release of insulin from the pancreas. This mechanism of hormone production is stimulated by:

a. humoral stimuli
b. hormonal stimuli
c. neural stimuli
d. negative stimuli

Solution:

A

Exercise:

Problem:

Which mechanism of hormonal stimulation would be affected if signaling and hormone release from the hypothalamus was blocked?

a. humoral and hormonal stimuli
b. hormonal and neural stimuli
c. neural and humoral stimuli
d. hormonal and negative stimuli

Solution:

B

Free Response

Exercise:

Problem:

How is hormone production and release primarily controlled?

Solution:

Hormone production and release are primarily controlled by negative feedback. In negative feedback systems, a stimulus causes the release of a substance whose effects then inhibit further release. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then feed back to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland.

Exercise:

Problem: Compare and contrast hormonal and humoral stimuli.

Solution:

The term humoral is derived from the term humor, which refers to bodily fluids such as blood. Humoral stimuli refer to the control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood. For example, a rise in blood glucose levels triggers the pancreatic release of insulin. Insulin causes blood glucose levels to drop, which signals the pancreas to stop producing insulin in a negative feedback loop.

Hormonal stimuli refer to the release of a hormone in response to another hormone. A number of endocrine glands release hormones when stimulated by hormones released by other endocrine organs. For example, the hypothalamus produces hormones that stimulate the anterior pituitary. The anterior pituitary in turn releases hormones that regulate hormone production by other endocrine glands. For example, the anterior pituitary releases thyroid-stimulating hormone, which stimulates the thyroid gland to produce the hormones T_3 and T_4. As

blood concentrations of T_3 and T_4 rise they inhibit both the pituitary and the hypothalamus in a negative feedback loop.

Glossary

hormonal stimuli
: release of a hormone in response to another hormone

humoral stimuli
: control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood

neural stimuli
: stimulation of endocrine glands by the nervous system

Derived copy of Endocrine Glands
By the end of this section, you will be able to:

- Describe the role of different glands in the endocrine system
- Explain how the different glands work together to maintain homeostasis

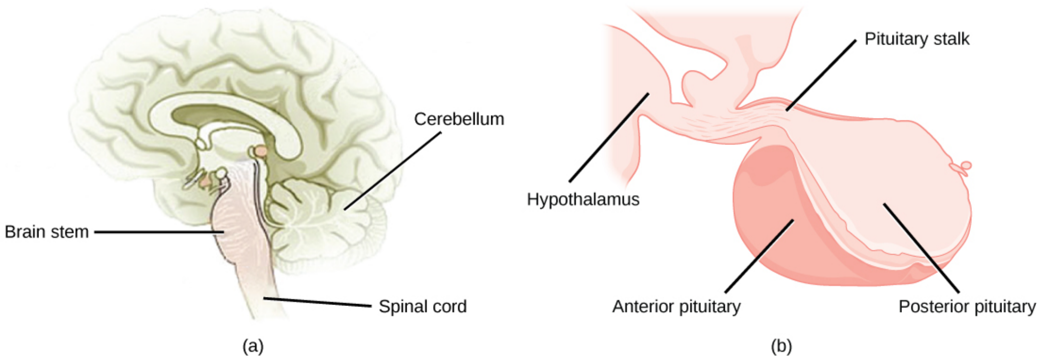
Both the endocrine and nervous systems use chemical signals to communicate and regulate the body's physiology. The endocrine system releases hormones that act on target cells to regulate development, growth, energy metabolism, reproduction, and many behaviors. The nervous system releases neurotransmitters or neurohormones that regulate neurons, muscle cells, and endocrine cells. Because the neurons can regulate the release of hormones, the nervous and endocrine systems work in a coordinated manner to regulate the body's physiology.

Hypothalamic-Pituitary Axis

The **hypothalamus** in vertebrates integrates the endocrine and nervous systems. The hypothalamus is an endocrine organ located in the diencephalon of the brain. It receives input from the body and other brain areas and initiates endocrine responses to environmental changes. The hypothalamus acts as an endocrine organ, synthesizing hormones and transporting them along axons to the posterior pituitary gland. It synthesizes and secretes regulatory hormones that control the endocrine cells in the anterior pituitary gland. The hypothalamus contains autonomic centers that control endocrine cells in the adrenal medulla via neuronal control.

The **pituitary gland**, sometimes called the hypophysis or "master gland" is located at the base of the brain in the sella turcica, a groove of the sphenoid bone of the skull, illustrated in [link]. It is attached to the hypothalamus via a stalk called the **pituitary stalk** (or infundibulum). The anterior portion of the pituitary gland is regulated by releasing or release-inhibiting hormones produced by the hypothalamus, and the posterior pituitary receives signals via neurosecretory cells to release hormones produced by the hypothalamus. The pituitary has two distinct regions—the anterior pituitary and the posterior pituitary—which between them secrete nine different

peptide or protein hormones. The posterior lobe of the pituitary gland contains axons of the hypothalamic neurons.



The pituitary gland is located at (a) the base of the brain and (b) connected to the hypothalamus by the pituitary stalk. (credit a: modification of work by NCI; credit b: modification of work by Gray's Anatomy)

Anterior Pituitary

The **anterior pituitary** gland, or adenohypophysis, is surrounded by a capillary network that extends from the hypothalamus, down along the infundibulum, and to the anterior pituitary. This capillary network is a part of the **hypophyseal portal system** that carries substances from the hypothalamus to the anterior pituitary and hormones from the anterior pituitary into the circulatory system. A portal system carries blood from one capillary network to another; therefore, the hypophyseal portal system allows hormones produced by the hypothalamus to be carried directly to the anterior pituitary without first entering the circulatory system.

The anterior pituitary produces seven hormones: growth hormone (GH), prolactin (PRL), thyroid-stimulating hormone (TSH), melanin-stimulating hormone (MSH), adrenocorticotropic hormone (ACTH), follicle-

stimulating hormone (FSH), and luteinizing hormone (LH). Anterior pituitary hormones are sometimes referred to as tropic hormones, because they control the functioning of other organs. While these hormones are produced by the anterior pituitary, their production is controlled by regulatory hormones produced by the hypothalamus. These regulatory hormones can be releasing hormones or inhibiting hormones, causing more or less of the anterior pituitary hormones to be secreted. These travel from the hypothalamus through the hypophyseal portal system to the anterior pituitary where they exert their effect. Negative feedback then regulates how much of these regulatory hormones are released and how much anterior pituitary hormone is secreted.

Posterior Pituitary

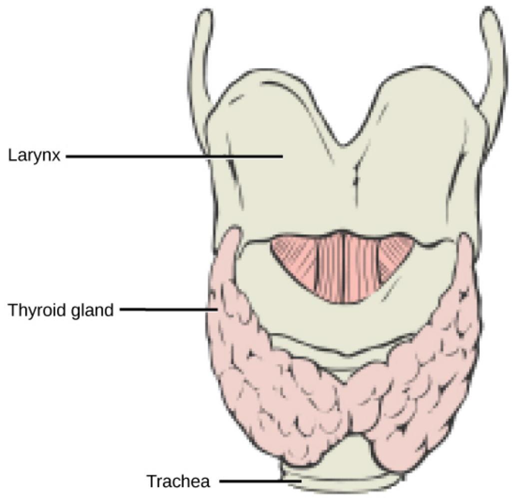
The **posterior pituitary** is significantly different in structure from the anterior pituitary. It is a part of the brain, extending down from the hypothalamus, and contains mostly nerve fibers and neuroglial cells, which support axons that extend from the hypothalamus to the posterior pituitary. The posterior pituitary and the infundibulum together are referred to as the neurohypophysis.

The hormones antidiuretic hormone (ADH), also known as vasopressin, and oxytocin are produced by neurons in the hypothalamus and transported within these axons along the infundibulum to the posterior pituitary. They are released into the circulatory system via neural signaling from the hypothalamus. These hormones are considered to be posterior pituitary hormones, even though they are produced by the hypothalamus, because that is where they are released into the circulatory system. The posterior pituitary itself does not produce hormones, but instead stores hormones produced by the hypothalamus and releases them into the blood stream.

Thyroid Gland

The **thyroid gland** is located in the neck, just below the larynx and in front of the trachea, as shown in [link]. It is a butterfly-shaped gland with two

lobes that are connected by the **isthmus**. It has a dark red color due to its extensive vascular system. When the thyroid swells due to dysfunction, it can be felt under the skin of the neck.



This illustration shows the location of the thyroid gland.

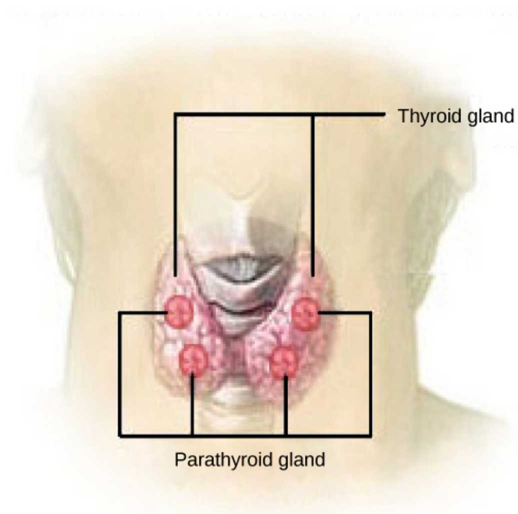
The thyroid gland is made up of many spherical thyroid follicles, which are lined with a simple cuboidal epithelium. These follicles contain a viscous fluid, called **colloid**, which stores the glycoprotein thyroglobulin, the precursor to the thyroid hormones. The follicles produce hormones that can be stored in the colloid or released into the surrounding capillary network for transport to the rest of the body via the circulatory system.

Thyroid follicle cells synthesize the hormone thyroxine, which is also known as T_4 because it contains four atoms of iodine, and triiodothyronine, also known as T_3 because it contains three atoms of iodine. Follicle cells are stimulated to release stored T_3 and T_4 by thyroid stimulating hormone (TSH), which is produced by the anterior pituitary. These thyroid hormones increase the rates of mitochondrial ATP production.

A third hormone, calcitonin, is produced by **parafollicular cells** of the thyroid either releasing hormones or inhibiting hormones. Calcitonin release is not controlled by TSH, but instead is released when calcium ion concentrations in the blood rise. Calcitonin functions to help regulate calcium concentrations in body fluids. It acts in the bones to inhibit osteoclast activity and in the kidneys to stimulate excretion of calcium. The combination of these two events lowers body fluid levels of calcium.

Parathyroid Glands

Most people have four **parathyroid glands**; however, the number can vary from two to six. These glands are located on the posterior surface of the thyroid gland, as shown in [link]. Normally, there is a superior gland and an inferior gland associated with each of the thyroid's two lobes. Each parathyroid gland is covered by connective tissue and contains many secretory cells that are associated with a capillary network.

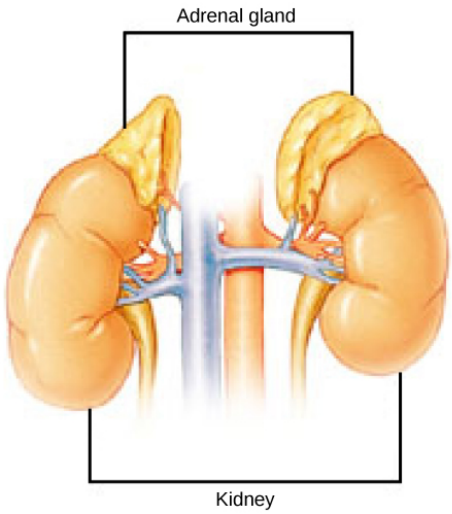


The parathyroid glands are located on the posterior of the thyroid gland. (credit: modification of work by NCI)

The parathyroid glands produce parathyroid hormone (PTH). PTH increases blood calcium concentrations when calcium ion levels fall below normal. PTH (1) enhances reabsorption of Ca^{2+} by the kidneys, (2) stimulates osteoclast activity and inhibits osteoblast activity, and (3) it stimulates synthesis and secretion of calcitriol by the kidneys, which enhances Ca^{2+} absorption by the digestive system. PTH is produced by chief cells of the parathyroid. PTH and calcitonin work in opposition to one another to maintain homeostatic Ca^{2+} levels in body fluids. Another type of cells, oxyphil cells, exist in the parathyroid but their function is not known. These hormones encourage bone growth, muscle mass, and blood cell formation in children and women.

Adrenal Glands

The **adrenal glands** are associated with the kidneys; one gland is located on top of each kidney as illustrated in [link]. The adrenal glands consist of an outer adrenal cortex and an inner adrenal medulla. These regions secrete different hormones.



The location of the adrenal glands on top of the kidneys

is shown. (credit: modification of work by NCI)

Adrenal Cortex

The **adrenal cortex** is made up of layers of epithelial cells and associated capillary networks. These layers form three distinct regions: an outer zona glomerulosa that produces mineralocorticoids, a middle zona fasciculata that produces glucocorticoids, and an inner zona reticularis that produces androgens.

The main mineralocorticoid is aldosterone, which regulates the concentration of Na^+ ions in urine, sweat, pancreas, and saliva. Aldosterone release from the adrenal cortex is stimulated by a decrease in blood concentrations of sodium ions, blood volume, or blood pressure, or by an increase in blood potassium levels.

The three main glucocorticoids are cortisol, corticosterone, and cortisone. The glucocorticoids stimulate the synthesis of glucose and gluconeogenesis (converting a non-carbohydrate to glucose) by liver cells and they promote the release of fatty acids from adipose tissue. These hormones increase blood glucose levels to maintain levels within a normal range between meals. These hormones are secreted in response to ACTH and levels are regulated by negative feedback.

Androgens are sex hormones that promote masculinity. They are produced in small amounts by the adrenal cortex in both males and females. They do not affect sexual characteristics and may supplement sex hormones released from the gonads.

Adrenal Medulla

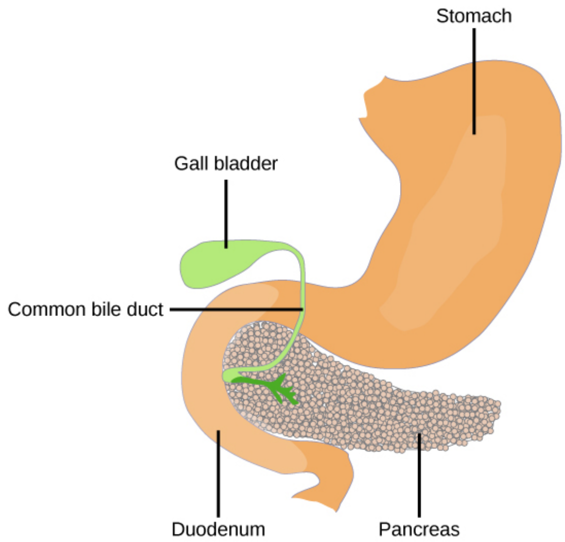
The **adrenal medulla** contains large, irregularly shaped cells that are closely associated with blood vessels. These cells are innervated by preganglionic autonomic nerve fibers from the central nervous system.

The adrenal medulla contains two types of secretory cells: one that produces epinephrine (adrenaline) and another that produces norepinephrine (noradrenaline). Epinephrine is the primary adrenal medulla hormone accounting for 75 to 80 percent of its secretions. Epinephrine and norepinephrine increase heart rate, breathing rate, cardiac muscle contractions, blood pressure, and blood glucose levels. They also accelerate the breakdown of glucose in skeletal muscles and stored fats in adipose tissue.

The release of epinephrine and norepinephrine is stimulated by neural impulses from the sympathetic nervous system. Secretion of these hormones is stimulated by acetylcholine release from preganglionic sympathetic fibers innervating the adrenal medulla. These neural impulses originate from the hypothalamus in response to stress to prepare the body for the fight-or-flight response.

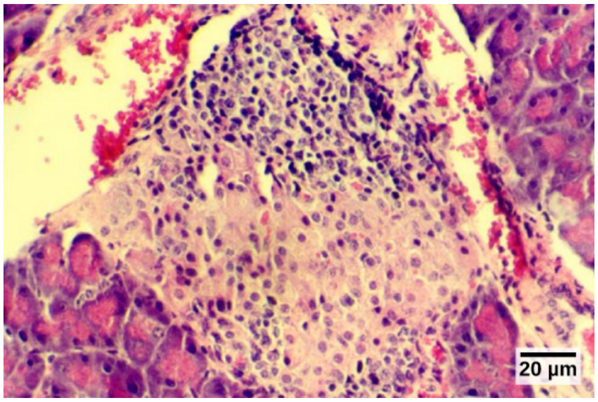
Pancreas

The **pancreas**, illustrated in [link], is an elongated organ that is located between the stomach and the proximal portion of the small intestine. It contains both exocrine cells that excrete digestive enzymes and endocrine cells that release hormones. It is sometimes referred to as a heterocrine gland because it has both endocrine and exocrine functions.



The pancreas is found underneath the stomach and points toward the spleen. (credit: modification of work by NCI)

The endocrine cells of the pancreas form clusters called pancreatic islets or the **islets of Langerhans**, as visible in the micrograph shown in [link]. The pancreatic islets contain two primary cell types: **alpha cells**, which produce the hormone glucagon, and **beta cells**, which produce the hormone insulin. These hormones regulate blood glucose levels. As blood glucose levels decline, alpha cells release glucagon to raise the blood glucose levels by increasing rates of glycogen breakdown and glucose release by the liver. When blood glucose levels rise, such as after a meal, beta cells release insulin to lower blood glucose levels by increasing the rate of glucose uptake in most body cells, and by increasing glycogen synthesis in skeletal muscles and the liver. Together, glucagon and insulin regulate blood glucose levels.



The islets of Langerhans are clusters of endocrine cells found in the pancreas; they stain lighter than surrounding cells. (credit: modification of work by Muhammad T. Tabiin, Christopher P. White, Grant Morahan, and Bernard E. Tuch; scale-bar data from Matt Russell)

Pineal Gland

The pineal gland produces melatonin. The rate of melatonin production is affected by the photoperiod. Collaterals from the visual pathways innervate the pineal gland. During the day photoperiod, little melatonin is produced; however, melatonin production increases during the dark photoperiod (night). In some mammals, melatonin has an inhibitory affect on reproductive functions by decreasing production and maturation of sperm, oocytes, and reproductive organs. Melatonin is an effective antioxidant, protecting the CNS from free radicals such as nitric oxide and hydrogen peroxide. Lastly, melatonin is involved in biological rhythms, particularly circadian rhythms such as the sleep-wake cycle and eating habits.

Gonads

The gonads—the male testes and female ovaries—produce steroid hormones. The testes produce androgens, testosterone being the most prominent, which allow for the development of secondary sex characteristics and the production of sperm cells. The ovaries produce estradiol and progesterone, which cause secondary sex characteristics and prepare the body for childbirth.

Endocrine Glands and their Associated Hormones		
Endocrine Gland	**Associated Hormones**	**Effect**
Hypothalamus	releasing and inhibiting hormones	regulate hormone release from pituitary gland; produce oxytocin; produce uterine contractions and milk secretion in females
	antidiuretic hormone (ADH)	water reabsorption from kidneys; vasoconstriction to increase blood pressure
Pituitary (Anterior)	growth hormone (GH)	promotes growth of body tissues, protein synthesis; metabolic functions
	prolactin (PRL)	promotes milk production

Endocrine Glands and their Associated Hormones		
Endocrine Gland	**Associated Hormones**	**Effect**
	thyroid stimulating hormone (TSH)	stimulates thyroid hormone release
	adrenocorticotropic hormone (ACTH)	stimulates hormone release by adrenal cortex, glucocorticoids
	follicle-stimulating hormone (FSH)	stimulates gamete production (both ova and sperm); secretion of estradiol
	luteinizing hormone (LH)	stimulates androgen production by gonads; ovulation, secretion of progesterone
	melanocyte-stimulating hormone (MSH)	stimulates melanocytes of the skin increasing melanin pigment production.
Pituitary (Posterior)	antidiuretic hormone (ADH)	stimulates water reabsorption by kidneys

Endocrine Glands and their Associated Hormones		
Endocrine Gland	**Associated Hormones**	**Effect**
	oxytocin	stimulates uterine contractions during childbirth; milk ejection; stimulates ductus deferens and prostate gland contraction during emission
Thyroid	thyroxine, triiodothyronine	stimulate and maintain metabolism; growth and development
	calcitonin	reduces blood Ca^{2+} levels
Parathyroid	parathyroid hormone (PTH)	increases blood Ca^{2+} levels
Adrenal (Cortex)	aldosterone	increases blood Na^{+} levels; increase K^{+} secretion
	cortisol, corticosterone, cortisone	increase blood glucose levels; anti-inflammatory effects
Adrenal (Medulla)	epinephrine, norepinephrine	stimulate fight-or-flight response; increase blood gluclose levels; increase metabolic activities

Endocrine Glands and their Associated Hormones		
Endocrine Gland	**Associated Hormones**	**Effect**
Pancreas	insulin	reduces blood glucose levels
	glucagon	increases blood glucose levels
Pineal gland	melatonin	regulates some biological rhythms and protects CNS from free radicals
Testes	androgens	regulate, promote, increase or maintain sperm production; male secondary sexual characteristics
Ovaries	estrogen	promotes uterine lining growth; female secondary sexual characteristics
	progestins	promote and maintain uterine lining growth

Section Summary

The pituitary gland is located at the base of the brain and is attached to the hypothalamus by the infundibulum. The anterior pituitary receives products from the hypothalamus by the hypophyseal portal system and produces six hormones. The posterior pituitary is an extension of the brain and releases

hormones (antidiuretic hormone and oxytocin) produced by the hypothalamus.

The thyroid gland is located in the neck and is composed of two lobes connected by the isthmus. The thyroid is made up of follicle cells that produce the hormones thyroxine and triiodothyronine. Parafollicular cells of the thyroid produce calcitonin. The parathyroid glands lie on the posterior surface of the thyroid gland and produce parathyroid hormone.

The adrenal glands are located on top of the kidneys and consist of the renal cortex and renal medulla. The adrenal cortex is the outer part of the adrenal gland and produces the corticosteroids, glucocorticoids, and mineralocorticoids. The adrenal medulla is the inner part of the adrenal gland and produces the catecholamines epinephrine and norepinephrine.

The pancreas lies in the abdomen between the stomach and the small intestine. Clusters of endocrine cells in the pancreas form the islets of Langerhans, which are composed of alpha cells that release glucagon and beta cells that release insulin.

Some organs possess endocrine activity as a secondary function but have another primary function. The heart produces the hormone atrial natriuretic peptide, which functions to reduce blood volume, pressure, and Na^+ concentration. The gastrointestinal tract produces various hormones that aid in digestion. The kidneys produce renin, calcitriol, and erythropoietin. Adipose tissue produces leptin, which promotes satiety signals in the brain.

Review Questions

Exercise:

Problem: Which endocrine glands are associated with the kidneys?

a. thyroid glands
b. pituitary glands
c. adrenal glands
d. gonads

Solution:

C

Exercise:

Problem:

Which of the following hormones is not produced by the anterior pituitary?

a. oxytocin
b. growth hormone
c. prolactin
d. thyroid-stimulating hormone

Solution:

A

Free Response

Exercise:

Problem: What does aldosterone regulate, and how is it stimulated?

Solution:

The main mineralocorticoid is aldosterone, which regulates the concentration of ions in urine, sweat, and saliva. Aldosterone release from the adrenal cortex is stimulated by a decrease in blood concentrations of sodium ions, blood volume, or blood pressure, or an increase in blood potassium levels.

Exercise:

Problem:

The adrenal medulla contains two types of secretory cells, what are they and what are their functions?

Solution:

The adrenal medulla contains two types of secretory cells, one that produces epinephrine (adrenaline) and another that produces norepinephrine (noradrenaline). Epinephrine is the primary adrenal medulla hormone accounting for 75–80 percent of its secretions. Epinephrine and norepinephrine increase heart rate, breathing rate, cardiac muscle contractions, and blood glucose levels. They also accelerate the breakdown of glucose in skeletal muscles and stored fats in adipose tissue. The release of epinephrine and norepinephrine is stimulated by neural impulses from the sympathetic nervous system. These neural impulses originate from the hypothalamus in response to stress to prepare the body for the fight-or-flight response.

Glossary

adrenal cortex
: outer portion of adrenal glands that produces corticosteroids

adrenal gland
: endocrine glands associated with the kidneys

adrenal medulla
: inner portion of adrenal glands that produces epinephrine and norepinephrine

alpha cell
: endocrine cell of the pancreatic islets that produces the hormone glucagon

anterior pituitary

portion of the pituitary gland that produces six hormones; also called adenohypophysis

atrial natriuretic peptide (ANP)
hormone produced by the heart to reduce blood volume, pressure, and Na^+ concentration

beta cell
endocrine cell of the pancreatic islets that produces the hormone insulin

colloid
fluid inside the thyroid gland that contains the glycoprotein thyroglobulin

endocrine gland
gland that secretes hormones into the surrounding interstitial fluid, which then diffuse into blood and are carried to various organs and tissues within the body

erythropoietin (EPO)
hormone produced by the kidneys to stimulate red blood cell production in the bone marrow

hypophyseal portal system
system of blood vessels that carries hormones from the hypothalamus to the anterior pituitary

islets of Langerhans (pancreatic islets)
endocrine cells of the pancreas

isthmus
tissue mass that connects the two lobes of the thyroid gland

leptin
hormone produced by adipose tissue that promotes feelings of satiety and reduces hunger

pancreas
: organ located between the stomach and the small intestine that contains exocrine and endocrine cells

parafollicular cell
: thyroid cell that produces the hormone calcitonin

parathyroid gland
: gland located on the surface of the thyroid that produces parathyroid hormone

pituitary gland
: endocrine gland located at the base of the brain composed of an anterior and posterior region; also called hypophysis

pituitary stalk
: (also, infundibulum) stalk that connects the pituitary gland to the hypothalamus

posterior pituitary
: extension of the brain that releases hormones produced by the hypothalamus; along with the infundibulum, it is also referred to as the neurohypophysis

thymus
: gland located behind the sternum that produces thymosin hormones that contribute to the development of the immune system

thyroid gland
: endocrine gland located in the neck that produces thyroid hormones thyroxine and triiodothyronine

Introduction
class="introduction"

Female seahorses produce eggs that are then fertilized by the male. Unlike with almost all other animals, the young then develop in a pouch of the male seahorse until birth. (credit: "cliff1066"/Flickr)

In the animal kingdom, each species has its unique adaptations for reproduction. Asexual reproduction produces genetically identical offspring (clones), whereas in sexual reproduction, the genetic material of two individuals combines to produce offspring that are genetically different from their parents. During sexual reproduction the male gamete (sperm) may be placed inside the female's body for internal fertilization, the sperm may be left in the environment for the female to pick up and place in her body, or both sperm and eggs may be released into the environment for external fertilization. Seahorses provide an example of the latter, but with a twist ([link]). Following a mating dance, the female releases eggs into the male seahorse's abdominal brood pouch and the male releases sperm into the water, which then find their way into the brood pouch to fertilize the eggs. The fertilized eggs develop in the pouch for several weeks.

How Animals Reproduce

By the end of this section, you will be able to:

- Describe advantages and disadvantages of asexual and sexual reproduction
- Discuss asexual reproduction methods
- Discuss sexual reproduction methods
- Discuss internal and external methods of fertilization

Some animals produce offspring through asexual reproduction while other animals produce offspring through sexual reproduction. Both methods have advantages and disadvantages. **Asexual reproduction** produces offspring that are genetically identical to the parent because the offspring are all clones of the original parent. A single individual can produce offspring asexually and large numbers of offspring can be produced quickly; these are two advantages that asexually reproducing organisms have over sexually reproducing organisms. In a stable or predictable environment, asexual reproduction is an effective means of reproduction because all the offspring will be adapted to that environment. In an unstable or unpredictable environment, species that reproduce asexually may be at a disadvantage because all the offspring are genetically identical and may not be adapted to different conditions.

During **sexual reproduction**, the genetic material of two individuals is combined to produce genetically diverse offspring that differ from their parents. The genetic diversity of sexually produced offspring is thought to give sexually reproducing individuals greater fitness because more of their offspring may survive and reproduce in an unpredictable or changing environment. Species that reproduce sexually (and have separate sexes) must maintain two different types of individuals, males and females. Only half the population (females) can produce the offspring, so fewer offspring will be produced when compared to asexual reproduction. This is a disadvantage of sexual reproduction compared to asexual reproduction.

Asexual Reproduction

Asexual reproduction occurs in prokaryotic microorganisms (bacteria and archaea) and in many eukaryotic, single-celled and multi-celled organisms. There are several ways that animals reproduce asexually, the details of which vary among individual species.

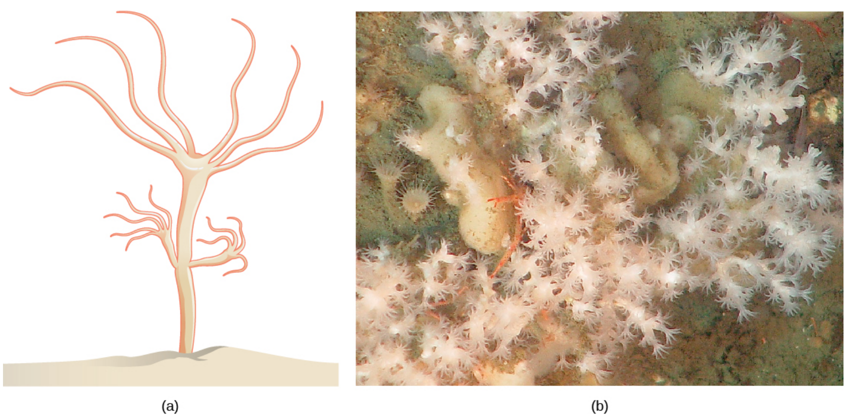
Fission

Fission, also called binary fission, occurs in some invertebrate, multi-celled organisms. It is in some ways analogous to the process of binary fission of single-celled prokaryotic organisms. The term fission is applied to instances in which an organism appears to split itself into two parts and, if necessary, regenerate the missing parts of each new organism. For example, species of turbellarian flatworms commonly called the planarians, such as *Dugesia dorotocephala*, are able to separate their bodies into head and tail regions and then regenerate the missing half in each of the two new organisms. Sea anemones (Cnidaria), such as species of the genus *Anthopleura* ([link]), will divide along the oral-aboral axis, and sea cucumbers (Echinodermata) of the genus *Holothuria*, will divide into two halves across the oral-aboral axis and regenerate the other half in each of the resulting individuals.

The *Anthopleura artemisia* sea anemone can reproduce through fission.

Budding

Budding is a form of asexual reproduction that results from the outgrowth of a part of the body leading to a separation of the “bud” from the original organism and the formation of two individuals, one smaller than the other. Budding occurs commonly in some invertebrate animals such as hydras and corals. In hydras, a bud forms that develops into an adult and breaks away from the main body ([link]).



(a) Hydra reproduce asexually through budding: a bud forms on the tubular body of an adult hydra, develops a mouth and tentacles, and then detaches from its parent. The new hydra is fully developed and will find its own location for attachment. (b) Some coral, such as the *Lophelia pertusa* shown here, can reproduce through budding. (credit b: modification of work by Ed Bowlby, NOAA/Olympic Coast NMS; NOAA/OAR/Office of Ocean Exploration)

Note:
Concept in Action

View this video to see a hydra budding.

Fragmentation

Fragmentation is the breaking of an individual into parts followed by regeneration. If the animal is capable of fragmentation, and the parts are big enough, a separate individual will regrow from each part. Fragmentation may occur through accidental damage, damage from predators, or as a natural form of reproduction. Reproduction through fragmentation is observed in sponges, some cnidarians, turbellarians, echinoderms, and annelids. In some sea stars, a new individual can be regenerated from a broken arm and a piece of the central disc. This sea star ([link]) is in the process of growing a complete sea star from an arm that has been cut off. Fisheries workers have been known to try to kill the sea stars eating their clam or oyster beds by cutting them in half and throwing them back into the ocean. Unfortunately for the workers, the two parts can each regenerate a new half, resulting in twice as many sea stars to prey upon the oysters and clams.

(a) *Linckia multifora* is a species of sea star that can reproduce asexually via fragmentation. In this process, (b) an arm that has been shed grows into a new sea star. (credit a: modifiction of work by Dwayne Meadows, NOAA/NMFS/OPR)

Parthenogenesis

Parthenogenesis is a form of asexual reproduction in which an egg develops into an individual without being fertilized. The resulting offspring can be either haploid or diploid, depending on the process in the species. Parthenogenesis occurs in invertebrates such as water fleas, rotifers, aphids, stick insects, and ants, wasps, and bees. Ants, bees, and wasps use parthenogenesis to produce haploid males (drones). The diploid females (workers and queens) are the result of a fertilized egg.

Some vertebrate animals—such as certain reptiles, amphibians, and fish—also reproduce through parthenogenesis. Parthenogenesis has been observed in species in which the sexes were separated in terrestrial or marine zoos. Two female Komodo dragons, a hammerhead shark, and a blacktop shark have produced parthenogenic young when the females have been isolated from males. It is possible that the asexual reproduction observed occurred in response to unusual circumstances and would normally not occur.

Sexual Reproduction

Sexual reproduction is the combination of reproductive cells from two individuals to form genetically unique offspring. The nature of the individuals that produce the two kinds of gametes can vary, having for example separate sexes or both sexes in each individual. Sex determination, the mechanism that determines which sex an individual develops into, also can vary.

Hermaphroditism

Hermaphroditism occurs in animals in which one individual has both male and female reproductive systems. Invertebrates such as earthworms, slugs, tapeworms, and snails ([link]) are often hermaphroditic. Hermaphrodites may self-fertilize, but typically they will mate with another of their species, fertilizing each other and both producing offspring. Self-fertilization is more common in animals that have limited mobility or are not motile, such as barnacles and clams. Many species have specific mechanisms in place to prevent self-fertilization, because it is an extreme form of inbreeding and usually produces less fit offspring.

(a) (b)

Many (a) snails are hermaphrodites. When two individuals (b) mate, they can produce up to 100 eggs each. (credit a: modification of work

by Assaf Shtilman; credit b: modification of work by "Schristia"/Flickr)

Sex Determination

Mammalian sex is determined genetically by the combination of X and Y chromosomes. Individuals homozygous for X (XX) are female and heterozygous individuals (XY) are male. In mammals, the presence of a Y chromosome causes the development of male characteristics and its absence results in female characteristics. The XY system is also found in some insects and plants.

Bird **sex determination** is dependent on the combination of Z and W chromosomes. Homozygous for Z (ZZ) results in a male and heterozygous (ZW) results in a female. Notice that this system is the opposite of the mammalian system because in birds the female is the sex with the different sex chromosomes. The W appears to be essential in determining the sex of the individual, similar to the Y chromosome in mammals. Some fish, crustaceans, insects (such as butterflies and moths), and reptiles use the ZW system.

More complicated chromosomal sex determining systems also exist. For example, some swordtail fish have three sex chromosomes in a population.

The sex of some other species is not determined by chromosomes, but by some aspect of the environment. Sex determination in alligators, some turtles, and tuataras, for example, is dependent on the temperature during the middle third of egg development. This is referred to as environmental sex determination, or more specifically, as temperature-dependent sex determination. In many turtles, cooler temperatures during egg incubation produce males and warm temperatures produce females, while in many other species of turtles, the reverse is true. In some crocodiles and some turtles, moderate temperatures produce males and both warm and cool temperatures produce females.

Individuals of some species change their sex during their lives, switching from one to the other. If the individual is female first, it is termed protogyny or “first female,” if it is male first, it is termed protandry or “first male.” Oysters are born male, grow in size, and become female and lay eggs. The wrasses, a family of reef fishes, are all sequential hermaphrodites. Some of these species live in closely coordinated schools with a dominant male and a large number of smaller females. If the male dies, a female increases in size, changes sex, and becomes the new dominant male.

Fertilization

The fusion of a sperm and an egg is a process called fertilization. This can occur either inside (**internal fertilization**) or outside (**external fertilization**) the body of the female. Humans provide an example of the former, whereas frog reproduction is an example of the latter.

External Fertilization

External fertilization usually occurs in aquatic environments where both eggs and sperm are released into the water. After the sperm reaches the egg, fertilization takes place. Most external fertilization happens during the process of spawning where one or several females release their eggs and the male(s) release sperm in the same area, at the same time. The spawning may be triggered by environmental signals, such as water temperature or the length of daylight. Nearly all fish spawn, as do crustaceans (such as crabs and shrimp), mollusks (such as oysters), squid, and echinoderms (such as sea urchins and sea cucumbers). Revise to "Frogs, corals, squid, and octopuses also spawn ([link]).

During sexual reproduction in toads, the male grasps the female from behind and externally fertilizes the eggs as they are deposited. (credit: Bernie Kohl)

Internal Fertilization

Internal fertilization occurs most often in terrestrial animals, although some aquatic animals also use this method. Internal fertilization may occur by the male directly depositing sperm in the female during mating. It may also occur by the male depositing sperm in the environment, usually in a protective structure, which a female picks up to deposit the sperm in her reproductive tract. There are three ways that offspring are produced following internal fertilization. In **oviparity**, fertilized eggs are laid outside the female's body and develop there, receiving nourishment from the yolk that is a part of the egg ([link]**a**). This occurs in some bony fish, some reptiles, a few cartilaginous fish, some amphibians, a few mammals, and all birds. Most non-avian reptiles and insects produce leathery eggs, while birds and some turtles produce eggs with high concentrations of calcium carbonate in the shell, making them hard. Chicken eggs are an example of a hard shell. The eggs of the egg-laying mammals such as the platypus and echidna are leathery.

In **ovoviparity**, fertilized eggs are retained in the female, and the embryo obtains its nourishment from the egg’s yolk. The eggs are retained in the female’s body until they hatch inside of her, or she lays the eggs right before they hatch. This process helps protect the eggs until hatching. This occurs in some bony fish (like the platyfish *Xiphophorus maculatus*, [link]**b**), some sharks, lizards, some snakes (garter snake *Thamnophis sirtalis*), some vipers, and some invertebrate animals (Madagascar hissing cockroach *Gromphadorhina portentosa*).

In **viviparity** the young are born alive. They obtain their nourishment from the female and are born in varying states of maturity. This occurs in most mammals ([link]**c**), some cartilaginous fish, and a few reptiles.

(a) (b) (c)

In (a) oviparity, young develop in eggs outside the female body, as with these *Harmonia axydridis* beetles hatching. Some aquatic animals, like this (b) pregnant *Xiphophorus maculatus* are ovoviparous, with the egg developing inside the female and nutrition supplied primarily from the yolk. In mammals, nutrition is supported by the placenta, as was the case with this (c) newborn squirrel. (credit b: modification of work by Gourami Watcher; credit c: modification of work by "audreyjm529"/Flickr)

Section Summary

Reproduction may be asexual when one individual produces genetically identical offspring, or sexual when the genetic material from two individuals is combined to produce genetically diverse offspring. Asexual reproduction in animals occurs through fission, budding, fragmentation, and parthenogenesis. Sexual reproduction may involve fertilization inside the body or in the external environment. A species may have separate sexes or combined sexes; when the sexes are combined they may be expressed at different times in the life cycle. The sex of an individual may be determined by various chromosomal systems or environmental factors such as temperature.

Sexual reproduction starts with the combination of a sperm and an egg in a process called fertilization. This can occur either outside the bodies or inside the female. The method of fertilization varies among animals. Some species release the egg and sperm into the environment, some species retain the egg and receive the sperm into the female body and then expel the developing embryo covered with shell, while still other species retain the developing offspring throughout the gestation period.

Review Questions

Exercise:

Problem: In which group is parthenogenesis a normal event?

a. chickens
b. bees
c. rabbits
d. sea stars

Solution:

B

Exercise:

Problem:

Genetically unique individuals are produced through ________.

a. sexual reproduction
b. parthenogenesis
c. budding
d. fragmentation

Solution:

A

Exercise:

Problem: External fertilization occurs in which type of environment?

a. aquatic
b. forested
c. savanna
d. steppe

Solution:

A

Free Response

Exercise:

Problem:

What might be a disadvantage to temperature-dependent sex determination?

Solution:

Temperatures can vary from year to year and an unusually cold or hot year might produce offspring all of one sex, making it hard for individuals to find mates.

Exercise:

Problem:

Compared to separate sexes and assuming self-fertilizing is not possible, what might be one advantage and one disadvantage to hermaphroditism?

Solution:

A possible advantage of hermaphroditism might be that anytime an individual of the same species is encountered a mating is possible, unlike separate sexes that must find an individual of the right sex to mate. (Also, every individual in a hermaphrodite population is able to produce offspring, which is not the case in populations with separate sexes.) A disadvantage might be that hermaphrodite populations are less efficient because they do not specialize in one sex or another, which means a hermaphrodite does not produce as many offspring through eggs or sperm as do species with separate sexes. (Other answers are possible.)

Glossary

asexual reproduction
: a mechanism that produces offspring that are genetically identical to the parent

budding
: a form of asexual reproduction that results from the outgrowth of a part of an organism leading to a separation from the original animal into two individuals

external fertilization

the fertilization of eggs by sperm outside an animal’s body, often during spawning

fission
: (also, binary fission) a form of asexual reproduction in which an organism splits into two separate organisms or two parts that regenerate the missing portions of the body

fragmentation
: the breaking of an organism into parts and the growth of a separate individual from each part

hermaphroditism
: the state of having both male and female reproductive structures within the same individual

internal fertilization
: the fertilization of eggs by sperm inside the body of the female

oviparity
: a process by which fertilized eggs are laid outside the female’s body and develop there, receiving nourishment from the yolk that is a part of the egg

ovoviparity
: a process by which fertilized eggs are retained within the female; the embryo obtains its nourishment from the egg’s yolk, and the young are fully developed when they are hatched

parthenogenesis
: a form of asexual reproduction in which an egg develops into a complete individual without being fertilized

sex determination
: the mechanism by which the sex of individuals in sexually reproducing organisms is initially established

sexual reproduction

a form of reproduction in which cells containing genetic material from two individuals combines to produce genetically unique offspring

viviparity

a process in which the young develop within the female and are born in a nonembryonic state

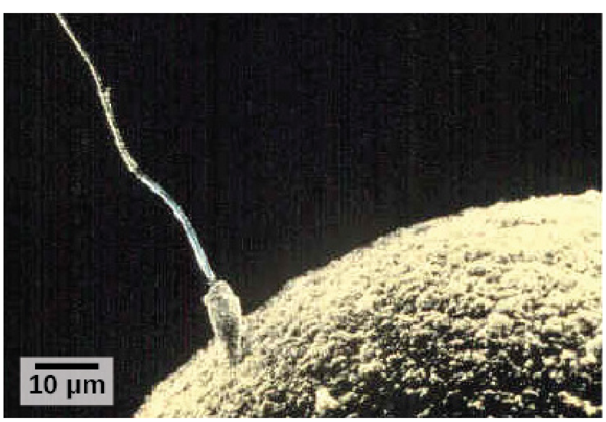
Development and Organogenesis
By the end of this section, you will be able to:

- Explain how the embryo forms from the zygote
- Discuss the role of cleavage and gastrulation in animal development
- Describe organogenesis

The process by which an organism develops from a single-celled zygote to a multi-cellular organism is complex and well regulated. The regulation occurs through signaling between cells and tissues and responses in the form of differential gene expression.

Early Embryonic Development

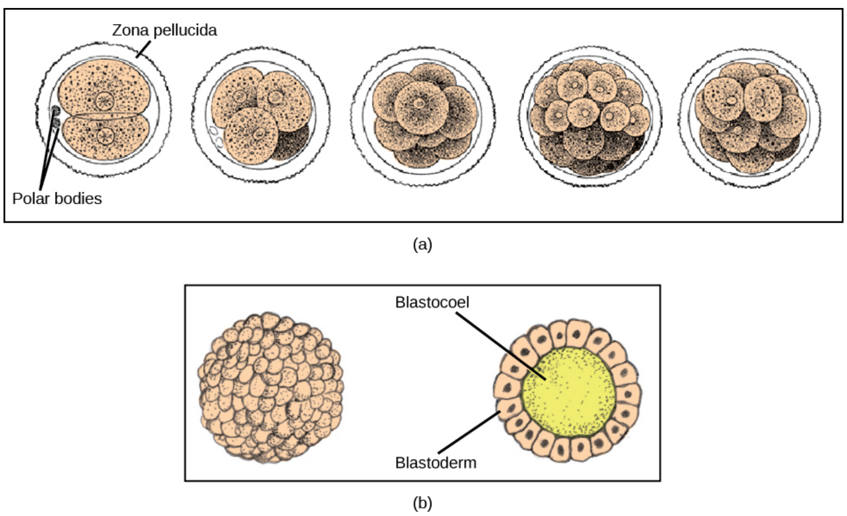
Fertilization is the process in which gametes (an egg and sperm) fuse to form a zygote ([link]). To ensure that the offspring has only one complete diploid set of chromosomes, only one sperm must fuse with one egg. In mammals, a layer called the **zona pellucida** protects the egg. At the tip of the head of a sperm cell is a structure like a lysosome called the acrosome, which contains enzymes. When a sperm binds to the zona pellucida, a series of events, called the acrosomal reactions, take place. These reactions, involving enzymes from the acrosome, allow the sperm plasma membrane to fuse with the egg plasma membrane and permit the sperm nucleus to transfer into the ovum. The nuclear membranes of the egg and sperm break down and the two haploid nuclei fuse to form a diploid nucleus or genome.



Fertilization is the process in which sperm and egg fuse to form a zygote. (credit: scale-bar data from Matt Russell)

To ensure that no more than one sperm fertilizes the egg, once the acrosomal reactions take place at one location of the egg membrane, the egg releases proteins in other locations to prevent other sperm from fusing with the egg.

The development of multi-cellular organisms begins from this single-celled zygote, which undergoes rapid cell division, called cleavage ([link]**a**), to form a hollow ball of cells called a blastula ([link]**b**).



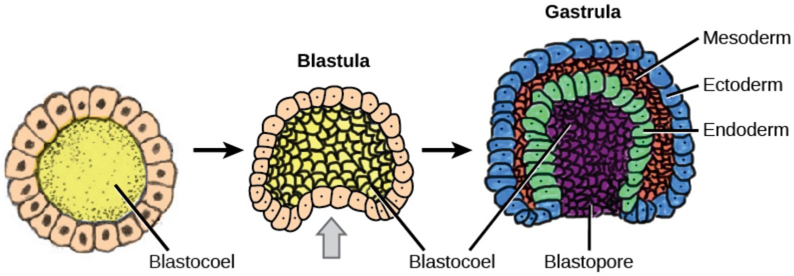
(a) During cleavage, the zygote rapidly divides into multiple cells. (b) The cells rearrange themselves to form a hollow ball called the blastula. (credit a: modification of work by Gray's Anatomy; credit b: modification of work by Pearson Scott Foresman; donated to the Wikimedia Foundation)

In mammals, the blastula forms the **blastocyst** in the next stage of development. Here the cells in the blastula arrange themselves in two layers: the **inner cell mass**, and an outer layer called the **trophoblast**. The inner cell mass will go on to form the embryo. The trophoblast secretes enzymes that allow implantation of the blastocyst into the endometrium of the uterus. The trophoblast will contribute to the placenta and nourish the embryo.

Note:
Concept in Action

Visit the Virtual Human Embryo project at the Endowment for Human Development site to click through an interactive of the stages of embryo development, including micrographs and rotating 3-D images.

The cells in the blastula then rearrange themselves spatially to form three layers of cells. This process is called **gastrulation**. During gastrulation, the blastula folds in on itself and cells migrate to form the three layers of cells ([link]) in a structure, the gastrula, with a hollow space that will become the digestive tract. Each of the layers of cells is called a germ layer and will differentiate into different organ systems.



Gastrulation is the process wherein the cells in the blastula rearrange themselves to form the germ layers. (credit: modification of work by Abigail Pyne)

The three germ layers are the endoderm, the ectoderm, and the mesoderm. Cells in each germ layer differentiate into tissues and embryonic organs. The ectoderm gives rise to the nervous system and the epidermis, among other tissues. The mesoderm gives rise to the muscle cells and connective tissue in the body. The endoderm gives rise to the gut and many internal organs.

Organogenesis

Gastrulation leads to the formation of the three germ layers that give rise during further development to the different organs in the animal body. This process is called **organogenesis**.

Organs develop from the germ layers through the process of differentiation. During differentiation, the embryonic stem cells express specific sets of genes that will determine their ultimate cell type. For example, some cells in the ectoderm will express the genes specific to skin cells. As a result, these cells will take on the shape and characteristics of epidermal cells. The process of differentiation is regulated by location-specific chemical signals from the cell’s embryonic environment that sets in play a cascade of events that regulates gene expression.

Section Summary

The early stages of embryonic development begin with fertilization. The process of fertilization is tightly controlled to ensure that only one sperm fuses with one egg. After fertilization, the zygote undergoes cleavage to form the blastula. The blastula, which in some species is a hollow ball of cells, undergoes a process called gastrulation, during which the three germ layers form. The ectoderm gives rise to the nervous system and the epidermal skin cells, the mesoderm gives rise to the muscle cells and connective tissue in the body, and the endoderm gives rise to the digestive system and other internal organs. Organogenesis is the formation of organs from the germ layers. Each germ layer gives rise to specific tissue types.

Review Questions

Exercise:

Problem: The process of gastrulation forms the ________.

a. blastula
b. zygote
c. organs
d. germ layers

Solution:

D

Exercise:

Problem: Which of the following gives rise to the skin cells?

a. ectoderm
b. endoderm
c. mesoderm
d. none of the above

Solution:

A

Free Response

Exercise:

Problem:

What do you think would happen if multiple sperm fused with one egg?

Solution:

If multiple sperm fused with one egg, a zygote with a multiple ploidy level (multiple copies of the chromosomes) would form, and then would die.

Glossary

blastocyst
: the structure formed when cells in the mammalian blastula separate into an inner and outer layer

gastrulation
: the process in which the blastula folds over itself to form the three germ layers

inner cell mass
: the inner layer of cells in the blastocyst, which becomes the embryo

organogenesis
: the process of organ formation during development

trophoblast

the outer layer of cells in the blastocyst, which gives rise to the embryo’s contribution to the placenta

zona pellucida
: the protective layer around the mammalian egg

Human Reproduction

By the end of this section, you will be able to:

- Describe human male and female reproductive anatomies
- Describe spermatogenesis and oogenesis and discuss their differences and similarities
- Describe the role of hormones in human reproduction
- Describe the roles of male and female reproductive hormones

As in all animals, the adaptations for reproduction in humans are complex. They involve specialized and different anatomies in the two sexes, a hormone regulation system, and specialized behaviors regulated by the brain and endocrine system.

Human Reproductive Anatomy

The reproductive tissues of male and female humans develop similarly *in utero* until about the seventh week of gestation when a low level of the hormone testosterone is released from the gonads of the developing male. Testosterone causes the primitive gonads to differentiate into male sexual organs. When testosterone is absent, the primitive gonads develop into ovaries. Tissues that produce a penis in males produce a clitoris in females. The tissue that will become the scrotum in a male becomes the labia in a female. Thus the male and female anatomies arise from a divergence in the development of what were once common embryonic structures.

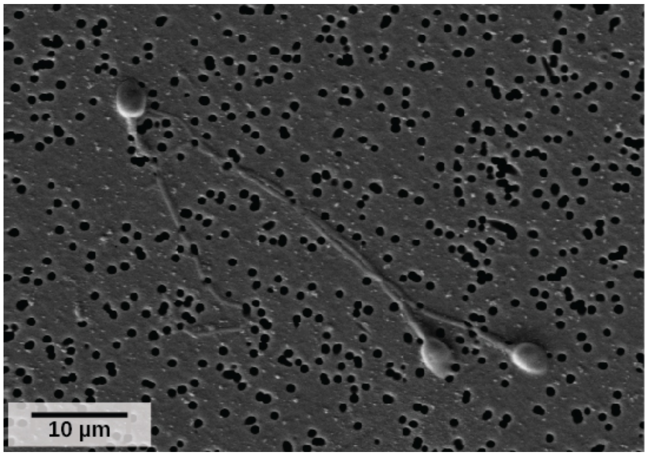
Male Reproductive Anatomy

Sperm are immobile at body temperature; therefore, the testes are external to the body so that a correct temperature is maintained for motility. In land mammals, including humans, the pair of testes must be suspended outside the body so the environment of the sperm is about 2 °C lower than body temperature to produce viable sperm. If the testes do not descend through the abdominal cavity during fetal development, the individual has reduced fertility.

The **scrotum** houses the testicles or **testes** (singular: testis), and provides passage for blood vessels, nerves, and muscles related to testicular function. The testes are a pair of male gonads that produce sperm and reproductive hormones. Each testis is approximately 2.5 by 3.8 cm (1.5 by 1 inch) in size and divided into wedge-shaped lobes by septa. Coiled in each wedge are seminiferous tubules that produce sperm.

The **penis** drains urine from the urinary bladder and is a copulatory organ during intercourse ([link]; [link]). The penis contains three tubes of erectile tissue that become engorged with blood, making the penis erect, in preparation for intercourse. The organ is inserted into the vagina culminating with an ejaculation. During orgasm, the accessory organs and glands connected to the testes contract and empty the semen (containing sperm) into the urethra and the fluid is expelled from the body by muscular contractions causing ejaculation. After intercourse, the blood drains from the erectile tissue and the penis becomes flaccid.

Semen is a mixture of sperm (about five percent of the total) and fluids from accessory glands that contribute most of the semen’s volume. Sperm are haploid cells, consisting of a flagellum for motility, a neck that contains the cell’s energy-producing mitochondria, and a head that contains the genetic material ([link]). An acrosome (acrosomal vesicle) is found at the top of the head of the sperm. This structure contains enzymes that can digest the protective coverings that surround the egg and allow the sperm to fuse with the egg. An ejaculate will contain from two to five milliliters of fluid and from 50–120 million sperm per milliliter.



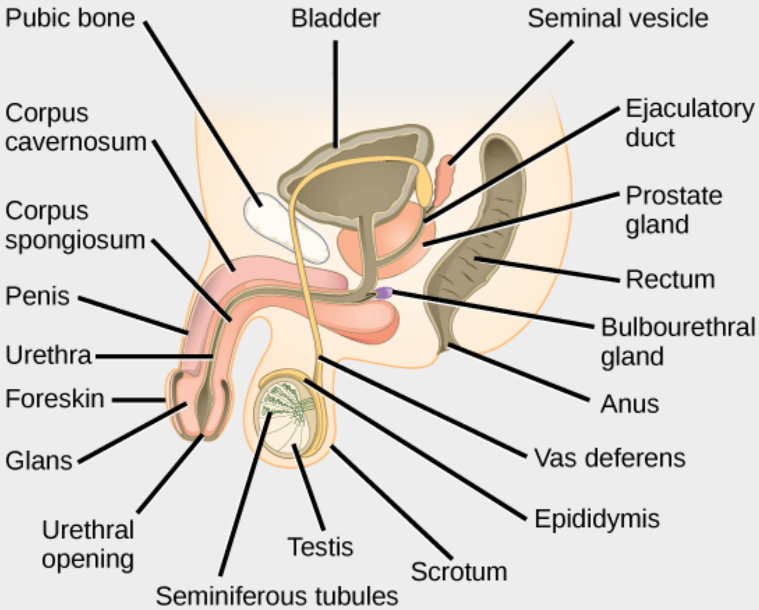
As seen in this scanning electron micrograph, human sperm has a flagellum, neck, and head. (credit: scale-bar data from Matt Russell)

Sperm form in the walls of **seminiferous tubules** that are coiled inside the testes ([link]; [link]). The walls of the seminiferous tubules are made up of the developing sperm cells, with the least developed sperm at the periphery of the tubule and the fully developed sperm next to the lumen. The sperm cells are associated with **Sertoli cells** that nourish and promote the development of the sperm. Other cells present between the walls of the tubules are the **interstitial cells of Leydig**, which produce testosterone once the male reaches adolescence.

When the sperm have developed flagella they leave the seminiferous tubules and enter the epididymis ([link]; [link]). This structure lies along the top and posterior of the testes and is the site of sperm maturation. The sperm leave the epididymis and enter the vas deferens, which carries the sperm behind the bladder, and forms the ejaculatory duct with the duct from the seminal vesicles. During a vasectomy, a section of the vas deferens is removed, preventing sperm (but not the secretions of the accessory glands) from being passed out of the body during ejaculation and preventing fertilization.

The bulk of the semen comes from the accessory glands associated with the male reproductive system. These are the **seminal vesicles**, the **prostate gland**, and the **bulbourethral gland** ([link]; [link]). The secretions from the accessory glands provide important compounds for the sperm including nutrients, electrolytes, and pH buffering. There are also coagulation factors that affect sperm delivery and motility.

Note:
Art Connection



The reproductive structures of the human male are shown.

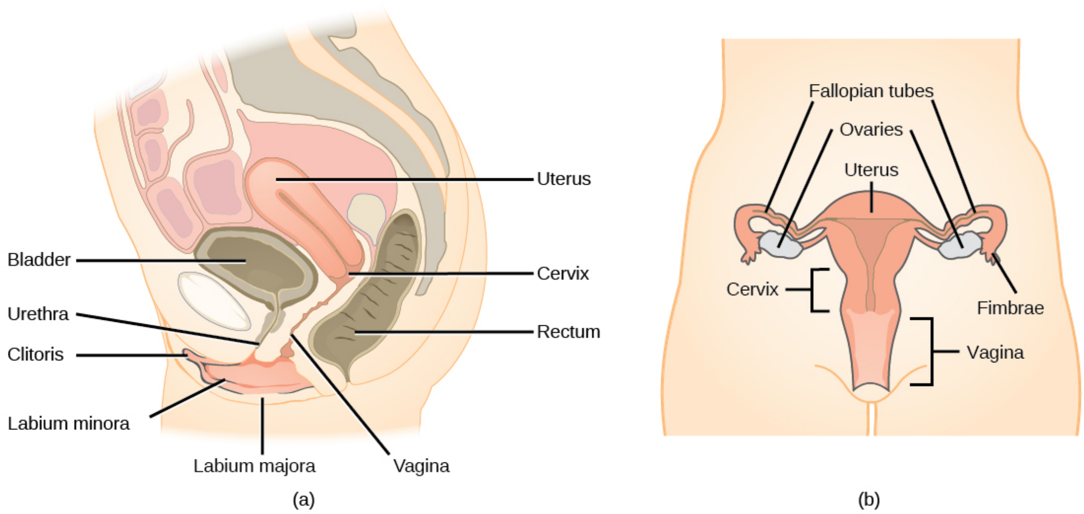
Which of the following statements about the male reproductive system is false?

a. The vas deferens carries sperm from the testes to the seminal vesicles.
b. The ejaculatory duct joins the urethra.
c. Both the prostate and the bulbourethral glands produce components of the semen.
d. The prostate gland is located in the testes.

Male Reproductive Anatomy		
Organ	**Location**	**Function**
Scrotum	External	Supports testes and regulates their temperature
Penis	External	Delivers urine, copulating organ
Testes	Internal	Produce sperm and male hormones
Seminal Vesicles	Internal	Contribute to semen production
Prostate Gland	Internal	Contributes to semen production
Bulbourethtral Glands	Internal	Neutralize urine in urethra

Female Reproductive Anatomy

A number of female reproductive structures are exterior to the body. These include the breasts and the vulva, which consists of the mons pubis, **clitoris**, **labia majora**, **labia minora**, and the vestibular glands ([link]; [link]).



The reproductive structures of the human female are shown. (credit a: modification of work by Gray's Anatomy; credit b: modification of work by CDC)

The breasts consist of mammary glands and fat. Each gland consists of 15 to 25 lobes that have ducts that empty at the nipple and that supply the nursing child with nutrient- and antibody-rich milk to aid development and protect the child.

Internal female reproductive structures include ovaries, oviducts, the uterus, and the vagina ([link]; [link]). The pair of ovaries is held in place in the abdominal cavity by a system of ligaments. The outermost layer of the ovary is made up of follicles, each consisting of one or more follicular cells that surround, nourish, and protect a single egg. During the menstrual period, a batch of follicular cells develops and prepares their eggs for release. At ovulation, one follicle ruptures and one egg is released. Following ovulation, the follicular tissue that surrounded the ovulated egg stays within the ovary and grows to form a solid mass called the **corpus luteum**. The corpus luteum secretes additional estrogen and the hormone progesterone that helps maintain the uterine lining during pregnancy. The ovaries also produce hormones, such as estrogen.

The **oviducts**, or fallopian tubes, extend from the uterus in the lower abdominal cavity to the ovaries, but they are not in contact with the ovaries. The lateral ends of the oviducts flare out into a trumpet-like structure and have a fringe of finger-like projections called fimbrae. When an egg is released at ovulation, the fimbrae help the nonmotile egg enter into the tube. The walls of the oviducts have a ciliated epithelium over smooth muscle. The cilia beat, and the smooth muscle contracts, moving the egg toward the uterus. Fertilization usually takes place within the oviduct and the developing embryo is moved toward the uterus. It usually takes the egg or embryo a week to travel through the oviduct.

Sterilization in women is called a tubal ligation; it is analogous to a vasectomy in males in that the oviducts are severed and sealed, preventing sperm from reaching the egg.

The **uterus** is a structure about the size of a woman’s fist. The uterus has a thick muscular wall and is lined with an endometrium rich in blood vessels and mucus glands that develop and thicken during the female cycle. Thickening of the endometrium prepares the uterus to receive the fertilized egg or zygote, which will then implant itself in the endometrium. The uterus supports the developing embryo and fetus during gestation. Contractions of the smooth muscle in the uterus aid in forcing the baby through the vagina during labor. If fertilization does not occur, a portion of the lining of the uterus sloughs off during each menstrual period. The endometrium builds up again in preparation for implantation. Part of the uterus, called the cervix, protrudes into the top of the vagina.

The **vagina** is a muscular tube that serves several purposes. It allows menstrual flow to leave the body. It is the receptacle for the penis during intercourse and the pathway for the delivery of offspring.

Female Reproductive Anatomy

Female Reproductive Anatomy

Organ	Location	Function
Clitoris	External	Sensory organ
Mons pubis	External	Fatty area overlying pubic bone
Labia majora	External	Covers labia minora; contains sweat and sebaceous glands
Labia minora	External	Covers vestibule
Greater vestibular glands	External	Secrete mucus; lubricate vagina
Breast	External	Produces and delivers milk
Ovaries	Internal	Produce and develop eggs
Oviducts	Internal	Transport egg to uterus; site of fertilization
Uterus	Internal	Supports developing embryo
Vagina	Internal	Common tube for intercourse, birth canal, passing menstrual flow

Gametogenesis (Spermatogenesis and Oogenesis)

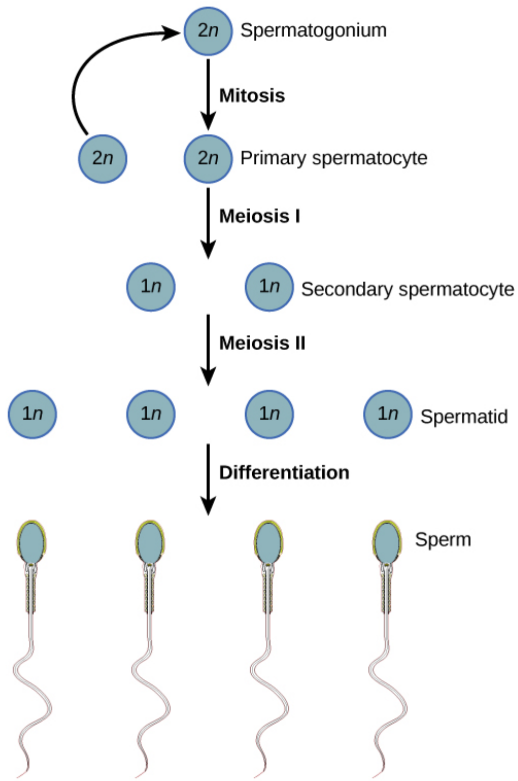
Gametogenesis, the production of sperm and eggs, involves the process of meiosis. During meiosis, two nuclear divisions separate the paired chromosomes in the nucleus and then separate the chromatids that were made during an earlier stage of the cell’s life cycle. Meiosis and its

associated cell divisions produces haploid cells with half of each pair of chromosomes normally found in diploid cells. The production of sperm is called **spermatogenesis** and the production of eggs is called **oogenesis**.

Spermatogenesis

Spermatogenesis occurs in the wall of the seminiferous tubules, with the most primitive cells at the periphery of the tube and the most mature sperm at the lumen of the tube ([link]). Immediately under the capsule of the tubule are diploid, undifferentiated cells. These stem cells, each called a spermatogonium (pl. spermatogonia), go through mitosis to produce one cell that remains as a stem cell and a second cell called a primary spermatocyte that will undergo meiosis to produce sperm.

The diploid primary spermatocyte goes through meiosis I to produce two haploid cells called secondary spermatocytes. Each secondary spermatocyte divides after meiosis II to produce two cells called spermatids. The spermatids eventually reach the lumen of the tubule and grow a flagellum, becoming sperm cells. Four sperm result from each primary spermatocyte that goes through meiosis.



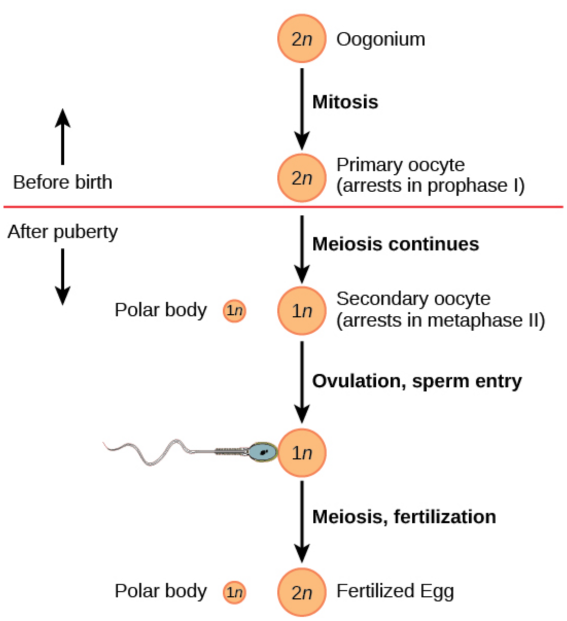
During spermatogenesis, four sperm result from each primary spermatocyte. The process also maps onto the physical structure of the wall of the seminiferous tubule, with the spermatogonia on the outer side of the tubule, and the sperm with their developing tails extended into the lumen of the tubule.

Note:
Concept in Action

Visit this site to see the process of spermatogenesis.

Oogenesis

Oogenesis occurs in the outermost layers of the ovaries. As with sperm production, oogenesis starts with a germ cell. In oogenesis, this germ cell is called an oogonium and forms during the embryological development of the individual. The oogonium undergoes mitosis to produce about one to two million oocytes by the time of birth.



The process of oogenesis occurs in the ovary’s outermost layer.

The primary oocytes begin meiosis before birth ([link]). However, the meiotic division is arrested in its progress in the first prophase stage. At the time of birth, all future eggs are in prophase I. This situation is in contrast with the male reproductive system in which sperm are produced continuously throughout the life of the individual. Starting at adolescence, anterior pituitary hormones cause the development of a few follicles in an ovary each month. This results in a primary oocyte finishing the first meiotic division. The cell divides unequally, with most of the cytoplasm and organelles going to one cell, called a secondary oocyte, and only one set of chromosomes and a small amount of cytoplasm going to the other cell. This second cell is called a polar body and usually dies. Cell division is again arrested, this time at metaphase II. At ovulation, this secondary oocyte is released and travels toward the uterus through the oviduct. If the secondary oocyte is fertilized, the cell continues through meiosis II, producing a second polar body and haploid egg, which fuses with the haploid sperm to form a fertilized egg (zygote) containing all 46 chromosomes.

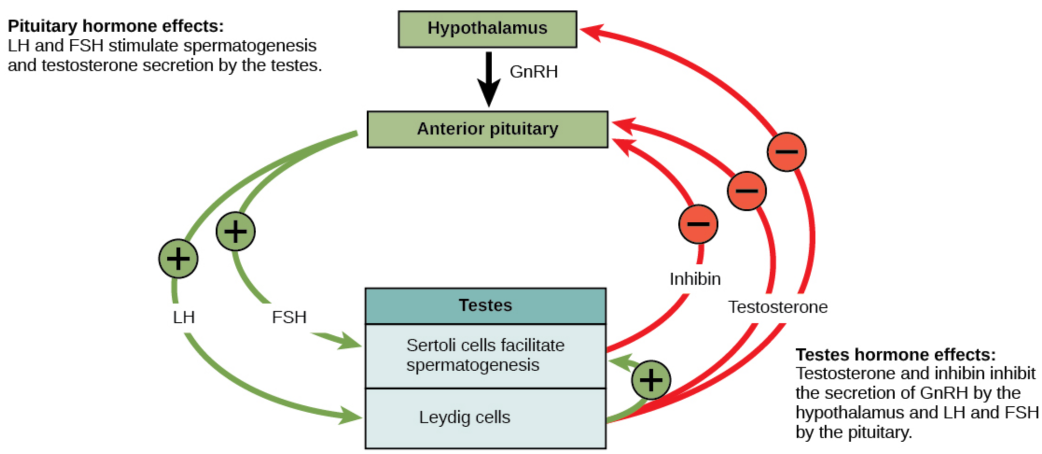
Hormonal Control of Reproduction

The human male and female reproductive cycles are controlled by the interaction of hormones from the hypothalamus and anterior pituitary with hormones from reproductive tissues and organs. In both sexes, the hypothalamus monitors and causes the release of hormones from the anterior pituitary gland. When the reproductive hormone is required, the hypothalamus sends a **gonadotropin-releasing hormone (GnRH)** to the anterior pituitary. This causes the release of **follicle stimulating hormone (FSH)** and **luteinizing hormone (LH)** from the anterior pituitary into the blood. Although these hormones are named after their functions in female reproduction, they are produced in both sexes and play important roles in controlling reproduction. Other hormones have specific functions in the male and female reproductive systems.

Male Hormones

At the onset of puberty, the hypothalamus causes the release of FSH and LH into the male system for the first time. FSH enters the testes and stimulates the Sertoli cells located in the walls of the seminiferous tubules to begin promoting spermatogenesis ([link]). LH also enters the testes and stimulates the interstitial cells of Leydig, located in between the walls of the seminiferous tubules, to make and release testosterone into the testes and the blood.

Testosterone stimulates spermatogenesis. This hormone is also responsible for the secondary sexual characteristics that develop in the male during adolescence. The secondary sex characteristics in males include a deepening of the voice, the growth of facial, axillary, and pubic hair, an increase in muscle bulk, and the beginnings of the sex drive.



Hormones control sperm production in a negative feedback system.

A negative feedback system occurs in the male with rising levels of testosterone acting on the hypothalamus and anterior pituitary to inhibit the release of GnRH, FSH, and LH. In addition, the Sertoli cells produce the hormone **inhibin**, which is released into the blood when the sperm count is

too high. This inhibits the release of GnRH and FSH, which will cause spermatogenesis to slow down. If the sperm count reaches a low of 20 million/mL, the Sertoli cells cease the release of inhibin, and the sperm count increases.

Female Hormones

The control of reproduction in females is more complex. The female reproductive cycle is divided into the ovarian cycle and the menstrual cycle. The **ovarian cycle** governs the preparation of endocrine tissues and release of eggs, while the **menstrual cycle** governs the preparation and maintenance of the uterine lining ([link]). These cycles are coordinated over a 22–32 day cycle, with an average length of 28 days.

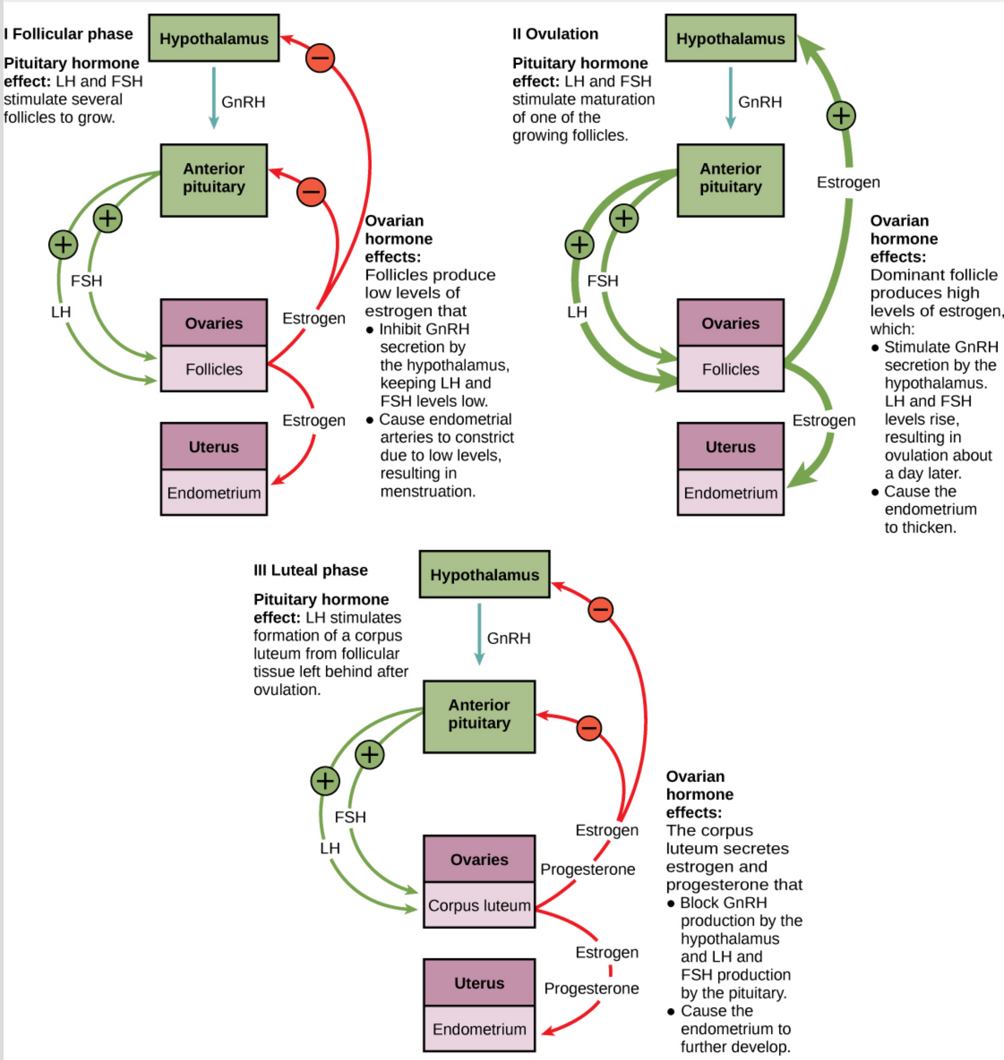
As with the male, the GnRH from the hypothalamus causes the release of the hormones FSH and LH from the anterior pituitary. In addition, **estrogen** and **progesterone** are released from the developing follicles. As with testosterone in males, estrogen is responsible for the secondary sexual characteristics of females. These include breast development, flaring of the hips, and a shorter period for bone growth.

The Ovarian Cycle and the Menstrual Cycle

The ovarian and menstrual cycles are regulated by hormones of the hypothalamus, pituitary, and ovaries ([link]). The ebb and flow of the hormones causes the ovarian and menstrual cycles to advance. The ovarian and menstrual cycles occur concurrently. The first half of the ovarian cycle is the follicular phase. Slowly rising levels of FSH cause the growth of follicles on the surface of the ovary. This process prepares the egg for ovulation. As the follicles grow, they begin releasing estrogen. The first few days of this cycle coincide with menstruation or the sloughing off of the functional layer of the endometrium in the uterus. After about five days, estrogen levels rise and the menstrual cycle enters the proliferative phase. The endometrium begins to regrow, replacing the blood vessels and glands that deteriorated during the end of the last cycle.

Note:

Art Connection



The ovarian and menstrual cycles of female reproduction are regulated by hormones produced by the hypothalamus, pituitary, and ovaries.

Which of the following statements about hormone regulation of the female reproductive cycle is false?

a. LH and FSH are produced in the pituitary, and estrogen and progesterone are produced in the ovaries.

b. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
c. Both progesterone and estrogen are produced by the follicles.
d. Secretion of GnRH by the hypothalamus is inhibited by low levels of estrogen but stimulated by high levels of estrogen.

Just prior to the middle of the cycle (approximately day 14), the high level of estrogen causes FSH and especially LH to rise rapidly then fall. The spike in LH causes the most mature follicle to rupture and release its egg. This is **ovulation**. The follicles that did not rupture degenerate and their eggs are lost. The level of estrogen decreases when the extra follicles degenerate.

Following ovulation, the ovarian cycle enters its luteal phase and the menstrual cycle enters its secretory phase, both of which run from about day 15 to 28. The luteal and secretory phases refer to changes in the ruptured follicle. The cells in the follicle undergo physical changes and produce a structure called a corpus luteum. The corpus luteum produces estrogen and progesterone. The progesterone facilitates the regrowth of the uterine lining and inhibits the release of further FSH and LH. The uterus is being prepared to accept a fertilized egg, should it occur during this cycle. The inhibition of FSH and LH prevents any further eggs and follicles from developing, while the progesterone is elevated. The level of estrogen produced by the corpus luteum increases to a steady level for the next few days.

If no fertilized egg is implanted into the uterus, the corpus luteum degenerates and the levels of estrogen and progesterone decrease. The endometrium begins to degenerate as the progesterone levels drop, initiating the next menstrual cycle. The decrease in progesterone also allows the hypothalamus to send GnRH to the anterior pituitary, releasing FSH and LH and starting the cycles again.

Note:
Career in Action
Reproductive Endocrinologist
A reproductive endocrinologist is a physician who treats a variety of hormonal disorders related to reproduction and infertility in both men and women. The disorders include menstrual problems, infertility, pregnancy loss, sexual dysfunction, and menopause. Doctors may use fertility drugs, surgery, or assisted reproductive techniques (ART) in their therapy. ART involves the use of procedures to manipulate the egg or sperm to facilitate reproduction, such as *in vitro* fertilization.
Reproductive endocrinologists undergo extensive medical training, first in a four-year residency in obstetrics and gynecology, then in a three-year fellowship in reproductive endocrinology. To be board certified in this area, the physician must pass written and oral exams in both areas.

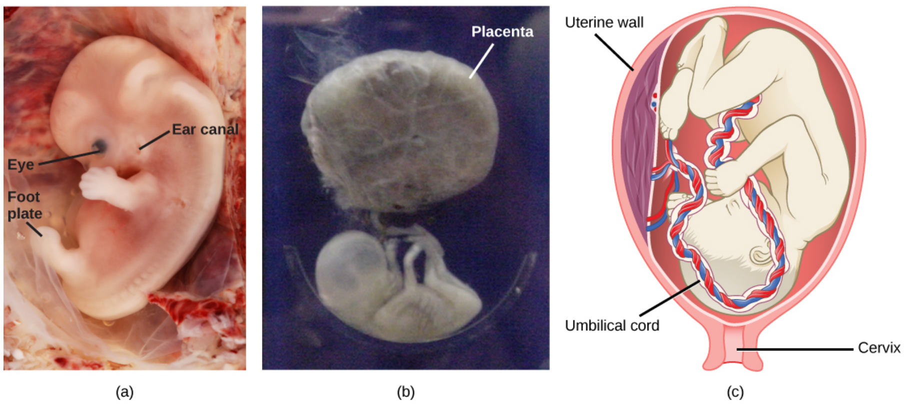
Gestation

Pregnancy begins with the fertilization of an egg and continues through to the birth of the individual. The length of time of **gestation**, or the **gestation period**, in humans is 266 days and is similar in other great apes.

Within 24 hours of fertilization, the egg nucleus has finished meiosis and the egg and sperm nuclei fuse. With fusion, the cell is known as a zygote. The zygote initiates cleavage and the developing embryo travels through the oviduct to the uterus. The developing embryo must implant into the wall of the uterus within seven days, or it will deteriorate and die. The outer layers of the developing embryo or blastocyst grow into the endometrium by digesting the endometrial cells, and healing of the endometrium closes up the blastocyst into the tissue. Another layer of the blastocyst, the chorion, begins releasing a hormone called **human beta chorionic gonadotropin (β-HCG)**, which makes its way to the corpus luteum and keeps that structure active. This ensures adequate levels of progesterone that will maintain the endometrium of the uterus for the support of the developing embryo. Pregnancy tests determine the level of β-HCG in urine or serum. If the hormone is present, the test is positive.

The gestation period is divided into three equal periods or trimesters. During the first two-to-four weeks of the first trimester, nutrition and waste are handled by the endometrial lining through diffusion. As the trimester progresses, the outer layer of the embryo begins to merge with the endometrium, and the placenta forms. The **placenta** takes over the nutrient and waste requirements of the embryo and fetus, with the mother's blood passing nutrients to the placenta and removing waste from it. Chemicals from the fetus, such as bilirubin, are processed by the mother's liver for elimination. Some of the mother's immunoglobulins will pass through the placenta, providing passive immunity against some potential infections.

Internal organs and body structures begin to develop during the first trimester. By five weeks, limb buds, eyes, the heart, and liver have been basically formed. By eight weeks, the term fetus applies, and the body is essentially formed ([link]**a**). The individual is about five centimeters (two inches) in length and many of the organs, such as the lungs and liver, are not yet functioning. Exposure to any toxins is especially dangerous during the first trimester, as all of the body's organs and structures are going through initial development. Anything that interferes with chemical signaling during that development can have a severe effect on the fetus' survival.



(a) Fetal development is shown at nine weeks gestation. (b) This fetus is just entering the second trimester, when the placenta takes over more of the functions performed as the baby develops. (c)

There is rapid fetal growth during the third trimester. (credit a: modification of work by Ed Uthman; credit b: modification of work by National Museum of Health and Medicine; credit c: modification of work by Gray's Anatomy)

During the second trimester, the fetus grows to about 30 cm (about 12 inches) ([link]**b**). It becomes active and the mother usually feels the first movements. All organs and structures continue to develop. The placenta has taken over the functions of nutrition and waste elimination and the production of estrogen and progesterone from the corpus luteum, which has degenerated. The placenta will continue functioning up through the delivery of the baby. During the third trimester, the fetus grows to 3 to 4 kg (6.5–8.5 lbs.) and about 50 cm (19–20 inches) long ([link]**c**). This is the period of the most rapid growth during the pregnancy as all organ systems continue to grow and develop.

Note:
Concept in Action

Visit this website to see the stages of human fetal development.

Labor is the muscular contractions to expel the fetus and placenta from the uterus. Toward the end of the third trimester, estrogen causes receptors on the uterine wall to develop and bind the hormone oxytocin. At this time, the baby reorients, facing forward and down with the back or crown of the head

engaging the cervix (uterine opening). This causes the cervix to stretch and nerve impulses are sent to the hypothalamus, which signals the release of oxytocin from the posterior pituitary. Oxytocin causes smooth muscle in the uterine wall to contract. At the same time, the placenta releases prostaglandins into the uterus, increasing the contractions. A positive feedback relay occurs between the uterus, hypothalamus, and the posterior pituitary to assure an adequate supply of oxytocin. As more smooth muscle cells are recruited, the contractions increase in intensity and force.

There are three stages to labor. During stage one, the cervix thins and dilates. This is necessary for the baby and placenta to be expelled during birth. The cervix will eventually dilate to about 10 cm. During stage two, the baby is expelled from the uterus. The uterus contracts and the mother pushes as she compresses her abdominal muscles to aid the delivery. The last stage is the passage of the placenta after the baby has been born and the organ has completely disengaged from the uterine wall. If labor should stop before stage two is reached, synthetic oxytocin, known as Pitocin, can be administered to restart and maintain labor.

Section Summary

The reproductive structures that evolved in land animals allow males and females to mate, fertilize internally, and support the growth and development of offspring. Gametogenesis, the production of sperm (spermatogenesis) and eggs (oogenesis), takes place through the process of meiosis.

The male and female reproductive cycles are controlled by hormones released from the hypothalamus and anterior pituitary and hormones from reproductive tissues and organs. The hypothalamus monitors the need for FSH and LH production and release from the anterior pituitary. FSH and LH affect reproductive structures to cause the formation of sperm and the preparation of eggs for release and possible fertilization. In the male, FSH and LH stimulate Sertoli cells and interstitial cells of Leydig in the testes to facilitate sperm production. The Leydig cells produce testosterone, which also is responsible for the secondary sexual characteristics of males. In females, FSH and LH cause estrogen and progesterone to be produced.

They regulate the female reproductive cycle, which is divided into the ovarian cycle and the menstrual cycle.

Human pregnancy begins with fertilization of an egg and proceeds through the three trimesters of gestation. The first trimester lays down the basic structures of the body, including the limb buds, heart, eyes, and the liver. The second trimester continues the development of all of the organs and systems. The third trimester exhibits the greatest growth of the fetus and culminates in labor and delivery. The labor process has three stages (contractions, delivery of the fetus, and expulsion of the placenta), each propelled by hormones.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the male reproductive system is false?

a. The vas deferens carries sperm from the testes to the seminal vesicles.
b. The ejaculatory duct joins the urethra.
c. Both the prostate and the bulbourethral glands produce components of the semen.
d. The prostate gland is located in the testes.

Solution:

[link] D

Exercise:

Problem:

[link] Which of the following statements about hormone regulation of the female reproductive cycle is false?

a. LH and FSH are produced in the pituitary, and estrogen and progesterone are produced in the ovaries.
b. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
c. Both progesterone and estrogen are produced by the follicles.
d. Secretion of GnRH by the hypothalamus is inhibited by low levels of estrogen but stimulated by high levels of estrogen.

Solution:

[link] C

Review Questions

Exercise:

Problem: Sperm are produced in the ________.

a. scrotum
b. seminal vesicles
c. seminiferous tubules
d. prostate gland

Solution:

C

Exercise:

Problem:

Which female organ has an endometrial lining that will support a developing baby?

a. labia minora
b. breast

c. ovaries
d. uterus

Solution:

D

Exercise:

Problem: Which hormone causes FSH and LH to be released?

a. testosterone
b. estrogen
c. GnRH
d. progesterone

Solution:

C

Exercise:

Problem:

Nutrient and waste requirements for the developing fetus are handled during the first few weeks by ________.

a. the placenta
b. diffusion through the endometrium
c. the chorion
d. the blastocyst

Solution:

B

Exercise:

Problem:

Which hormone is primarily responsible for the contractions during labor?

a. oxytocin
b. estrogen
c. β-HCG
d. progesterone

Solution:

A

Free Response

Exercise:

Problem:

Compare spermatogenesis and oogenesis as to timing of the processes, and the number and type of cells finally produced.

Solution:

Stem cells are laid down in the male during gestation and lie dormant until adolescence. Stem cells in the female increase to one to two million and enter the first meiotic division and are arrested in prophase. At adolescence, spermatogenesis begins and continues until death, producing the maximum number of sperm with each meiotic division. Oogenesis continues again at adolescence in batches of eggs with each menstrual cycle. These primary oocytes finish the first meiotic division, producing a viable egg with most of the cytoplasm and its contents, and a second cell called a polar body containing 23 chromosomes. The second meiotic division is initiated and arrested in

metaphase. At ovulation, one egg is released. If this egg is fertilized, it finishes the second meiotic division. This is a diploid, fertilized egg.

Exercise:

Problem:

Describe the events in the ovarian cycle leading up to ovulation.

Solution:

Low levels of progesterone allow the hypothalamus to send GnRH to the anterior pituitary and cause the release of FSH and LH. FSH stimulates follicles on the ovary to grow and prepare the eggs for ovulation. As the follicles increase in size, they begin to release estrogen and a low level of progesterone into the blood. The level of estrogen rises to a peak, causing a spike in the concentration of LH. This causes the most mature follicle to rupture and ovulation occurs.

Exercise:

Problem: Describe the stages of labor.

Solution:

Stage one of labor results in uterine contractions, which thin the cervix and dilate the cervical opening. Stage two delivers the baby, and stage three delivers the placenta.

Glossary

bulbourethral gland
: the paired glands in the human male that produce a secretion that cleanses the urethra prior to ejaculation

corpus luteum
: the endocrine tissue that develops from an ovarian follicle after ovulation; secretes progesterone and estrogen during pregnancy

clitoris

a sensory and erectile structure in female mammals, homologous to the male penis, stimulated during sexual arousal

estrogen

a reproductive hormone in females that assists in endometrial regrowth, ovulation, and calcium absorption

follicle stimulating hormone (FSH)

a reproductive hormone that causes sperm production in men and follicle development in women

gestation

the development before birth of a viviparous animal

gestation period

the length of time of development, from conception to birth, of the young of a viviparous animal

gonadotropin-releasing hormone (GnRH)

a hormone from the hypothalamus that causes the release of FSH and LH from the anterior pituitary

human beta chorionic gonadotropin (β-HCG)

a hormone produced by the chorion of the zygote that helps to maintain the corpus luteum and elevated levels of progesterone

inhibin

a hormone made by Sertoli cells, provides negative feedback to hypothalamus in control of FSH and GnRH release

interstitial cell of Leydig

a cell type found next to the seminiferous tubules that makes testosterone

labia majora

the large folds of tissue covering inguinal area

labia minora
the smaller folds of tissue within labia majora

luteinizing hormone (LH)
a reproductive hormone in both men and women, causes testosterone production in men and ovulation and lactation in women

menstrual cycle
the cycle of the degradation and re-growth of the endometrium

oogenesis
the process of producing haploid eggs

ovarian cycle
the cycle of preparation of egg for ovulation and the conversion of the follicle to the corpus luteum

oviduct
(also, fallopian tube) the muscular tube connecting uterus with ovary area

ovulation
the release of an oocyte from a mature follicle in the ovary of a vertebrate

penis
the male reproductive structure for urine elimination and copulation

placenta
the organ that supports the transport of nutrients and waste between the mothers and fetus’ blood in eutherian mammals

progesterone
a reproductive hormone in women; assists in endometrial regrowth and inhibition of FSH and LH release

prostate gland

a structure that is a mixture of smooth muscle and glandular material and that contributes to semen

scrotum
: a sac containing testes, exterior to body

semen
: a fluid mixture of sperm and supporting materials

seminal vesicle
: a secretory accessory gland in male; contributes to semen

seminiferous tubule
: the structures within which sperm production occurs in the testes

Sertoli cell
: a cell in the walls of the seminiferous tubules that assists developing sperm and secretes inhibin

spermatogenesis
: the process of producing haploid sperm

testes
: a pair of male reproductive organs

testosterone
: a reproductive hormone in men that assists in sperm production and promoting secondary sexual characteristics

uterus
: a female reproductive structure in which an embryo develops

vagina
: a muscular tube for the passage of menstrual flow, copulation, and birth of offspring

Introduction
class="introduction"

Asian carp jump out of the water in response to electrofishing. The Asian carp in the inset photograph were harvested from the Little Calumet River in Illinois in May, 2010, using rotenone, a toxin often used as an insecticide, in an effort to learn more about the population of the species. (credit main image: modification of work by USGS; credit inset: modification

of work by
Lt. David
French,
USCG)

Imagine sailing down a river in a small motorboat on a weekend afternoon; the water is smooth, and you are enjoying the sunshine and cool breeze when suddenly you are hit in the head by a 20-pound silver carp. This is a risk now on many rivers and canal systems in Illinois and Missouri because of the presence of Asian carp.

This fish—actually a group of species including the silver, black, grass, and big head carp—has been farmed and eaten in China for over 1,000 years. It is one of the most important aquaculture food resources worldwide. In the United States, however, Asian carp is considered a dangerous invasive species that disrupts ecological community structure to the point of threatening native species.

The effects of invasive species (such as the Asian carp, kudzu vine, predatory snakehead fish, and zebra mussel) are just one aspect of what ecologists study to understand how populations interact within ecological

communities, and what impact natural and human-induced disturbances have on the characteristics of communities.

Population Demographics and Dynamics
By the end of this section, you will be able to:

- Describe how ecologists measure population size and density
- Describe three different patterns of population distribution
- Use life tables to calculate mortality rates
- Describe the three types of survivorship curves and relate them to specific populations

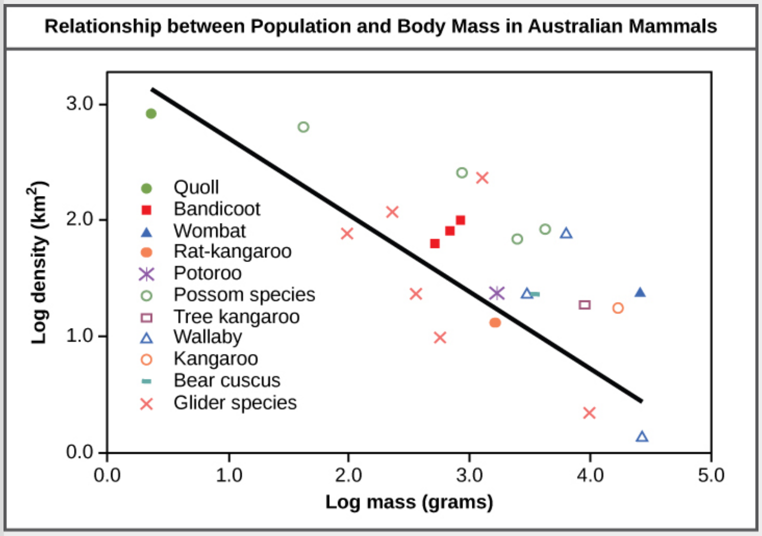
Populations are dynamic entities. Their size and composition fluctuate in response to numerous factors, including seasonal and yearly changes in the environment, natural disasters such as forest fires and volcanic eruptions, and competition for resources between and within species. The statistical study of populations is called **demography**: a set of mathematical tools designed to describe populations and investigate how they change. Many of these tools were actually designed to study human populations. For example, **life tables**, which detail the life expectancy of individuals within a population, were initially developed by life insurance companies to set insurance rates. In fact, while the term “demographics” is sometimes assumed to mean a study of human populations, all living populations can be studied using this approach.

Population Size and Density

Populations are characterized by their **population size** (total number of individuals) and their **population density** (number of individuals per unit area). A population may have a large number of individuals that are distributed densely, or sparsely. There are also populations with small numbers of individuals that may be dense or very sparsely distributed in a local area. Population size can affect potential for adaptation because it affects the amount of genetic variation present in the population. Density can have effects on interactions within a population such as competition for food and the ability of individuals to find a mate. Smaller organisms tend to be more densely distributed than larger organisms ([link]).

Note:

Art Connection



Australian mammals show a typical inverse relationship between population density and body size.

As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

Estimating Population Size

The most accurate way to determine population size is to count all of the individuals within the area. However, this method is usually not logistically or economically feasible, especially when studying large areas. Thus, scientists usually study populations by sampling a representative portion of each habitat and use this sample to make inferences about the population as a whole. The methods used to sample populations to determine their size and density are typically tailored to the characteristics of the organism being studied. For immobile organisms such as plants, or for very small and slow-moving organisms, a quadrat may be used. A **quadrat** is a wood, plastic, or metal square that is randomly located on the ground and used to count the number of individuals that lie within its boundaries. To obtain an accurate count using this method, the square must be placed at random locations

within the habitat enough times to produce an accurate estimate. This counting method will provide an estimate of both population size and density. The number and size of quadrat samples depends on the type of organisms and the nature of their distribution.

For smaller mobile organisms, such as mammals, a technique called **mark and recapture** is often used. This method involves marking a sample of captured animals in some way and releasing them back into the environment to mix with the rest of the population; then, a new sample is captured and scientists determine how many of the marked animals are in the new sample. This method assumes that the larger the population, the lower the percentage of marked organisms that will be recaptured since they will have mixed with more unmarked individuals. For example, if 80 field mice are captured, marked, and released into the forest, then a second trapping 100 field mice are captured and 20 of them are marked, the population size (*N*) can be determined using the following equation:

Equation:

Using our example, the population size would be 400.

Equation:

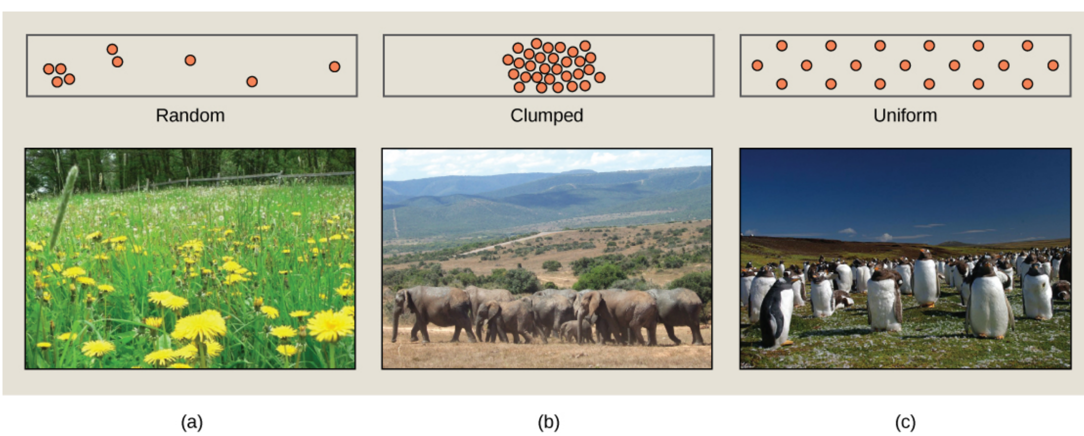
These results give us an estimate of 400 total individuals in the original population. The true number usually will be a bit different from this because of chance errors and possible bias caused by the sampling methods.

Species Distribution

In addition to measuring density, further information about a population can be obtained by looking at the distribution of the individuals throughout their range. A **species distribution pattern** is the distribution of individuals

within a habitat at a particular point in time—broad categories of patterns are used to describe them.

Individuals within a population can be distributed at random, in groups, or equally spaced apart (more or less). These are known as random, clumped, and uniform distribution patterns, respectively ([link]). Different distributions reflect important aspects of the biology of the species; they also affect the mathematical methods required to estimate population sizes. An example of random distribution occurs with dandelion and other plants that have wind-dispersed seeds that germinate wherever they happen to fall in favorable environments. A clumped distribution, may be seen in plants that drop their seeds straight to the ground, such as oak trees; it can also be seen in animals that live in social groups (schools of fish or herds of elephants). Uniform distribution is observed in plants that secrete substances inhibiting the growth of nearby individuals (such as the release of toxic chemicals by sage plants). It is also seen in territorial animal species, such as penguins that maintain a defined territory for nesting. The territorial defensive behaviors of each individual create a regular pattern of distribution of similar-sized territories and individuals within those territories. Thus, the distribution of the individuals within a population provides more information about how they interact with each other than does a simple density measurement. Just as lower density species might have more difficulty finding a mate, solitary species with a random distribution might have a similar difficulty when compared to social species clumped together in groups.



Species may have a random, clumped, or uniform distribution. Plants such as (a) dandelions with wind-dispersed seeds tend to be randomly distributed. Animals such as (b) elephants that travel in groups exhibit a clumped distribution. Territorial birds such as (c) penguins tend to have a uniform distribution. (credit a: modification of work by Rosendahl; credit b: modification of work by Rebecca Wood; credit c: modification of work by Ben Tubby)

Demography

While population size and density describe a population at one particular point in time, scientists must use demography to study the dynamics of a population. Demography is the statistical study of population changes over time: birth rates, death rates, and life expectancies. These population characteristics are often displayed in a life table.

Life Tables

Life tables provide important information about the life history of an organism and the life expectancy of individuals at each age. They are modeled after actuarial tables used by the insurance industry for estimating human life expectancy. Life tables may include the probability of each age group dying before their next birthday, the percentage of surviving individuals dying at a particular age interval (their **mortality rate**, and their life expectancy at each interval. An example of a life table is shown in [link] from a study of Dall mountain sheep, a species native to northwestern North America. Notice that the population is divided into age intervals (column A). The mortality rate (per 1000) shown in column D is based on the number of individuals dying during the age interval (column B), divided by the number of individuals surviving at the beginning of the interval (Column C) multiplied by 1000.

Equation:

For example, between ages three and four, 12 individuals die out of the 776 that were remaining from the original 1000 sheep. This number is then multiplied by 1000 to give the mortality rate per thousand.

Equation:

As can be seen from the mortality rate data (column D), a high death rate occurred when the sheep were between six months and a year old, and then increased even more from 8 to 12 years old, after which there were few survivors. The data indicate that if a sheep in this population were to survive to age one, it could be expected to live another 7.7 years on average, as shown by the life-expectancy numbers in column E.

Life Table of Dall Mountain Sheep[footnote] **Data Adapted from Edward S. Deevey, Jr., "Life Tables for Natural Populations of Animals," *The Quarterly Review of Biology* 22, no. 4 (December 1947): 283-314.**				
A	**B**	**C**	**D**	**E**

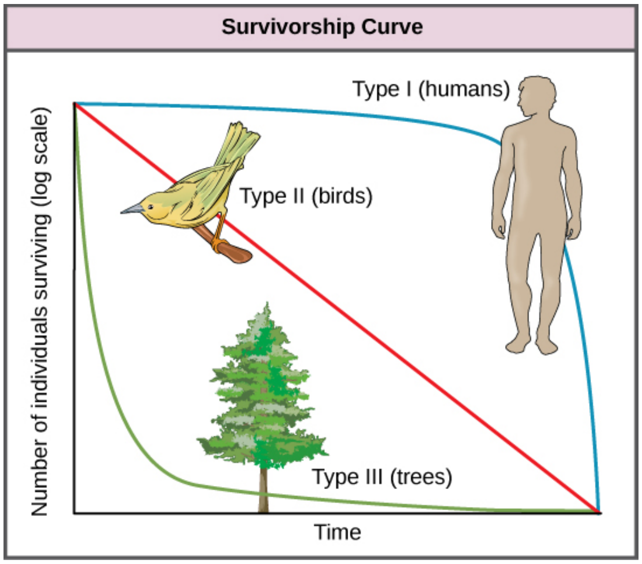
Age interval (years)	Number dying in age interval out of 1000 born	Number surviving at beginning of age interval out of 1000 born	Mortality rate per 1000 alive at beginning of age interval	Life expectancy or mean lifetime remaining to those attaining age interval
0–0.5	54	1000	54.0	7.06
0.5–1	145	946	153.3	—
1–2	12	801	15.0	7.7
2–3	13	789	16.5	6.8
3–4	12	776	15.5	5.9
4–5	30	764	39.3	5.0
5–6	46	734	62.7	4.2
6–7	48	688	69.8	3.4
7–8	69	640	107.8	2.6
8–9	132	571	231.2	1.9
9–10	187	439	426.0	1.3
10–11	156	252	619.0	0.9
11–12	90	96	937.5	0.6

12–13	3	6	500.0	1.2
13–14	3	3	1000	0.7

This life table of *Ovis dalli* shows the number of deaths, number of survivors, mortality rate, and life expectancy at each age interval for Dall mountain sheep.

Survivorship Curves

Another tool used by population ecologists is a **survivorship curve**, which is a graph of the number of individuals surviving at each age interval versus time. These curves allow us to compare the life histories of different populations ([link]). There are three types of survivorship curves. In a type I curve, mortality is low in the early and middle years and occurs mostly in older individuals. Organisms exhibiting a type I survivorship typically produce few offspring and provide good care to the offspring increasing the likelihood of their survival. Humans and most mammals exhibit a type I survivorship curve. In type II curves, mortality is relatively constant throughout the entire life span, and mortality is equally likely to occur at any point in the life span. Many bird populations provide examples of an intermediate or type II survivorship curve. In type III survivorship curves, early ages experience the highest mortality with much lower mortality rates for organisms that make it to advanced years. Type III organisms typically produce large numbers of offspring, but provide very little or no care for them. Trees and marine invertebrates exhibit a type III survivorship curve because very few of these organisms survive their younger years, but those that do make it to an old age are more likely to survive for a relatively long period of time.



Survivorship curves show the distribution of individuals in a population according to age. Humans and most mammals have a Type I survivorship curve, because death primarily occurs in the older years. Birds have a Type II survivorship curve, as death at any age is equally probable. Trees have a Type III survivorship curve because very few survive the younger years, but after a certain age, individuals are much more likely to survive.

Section Summary

Populations are individuals of a species that live in a particular habitat. Ecologists measure characteristics of populations: size, density, and distribution pattern. Life tables are useful to calculate life expectancies of individual population members. Survivorship curves show the number of individuals surviving at each age interval plotted versus time.

Exercise:

Problem:

[link] As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

Solution:

[link] Smaller animals require less food and others resources, so the environment can support more of them per unit area.

Multiple Choice

Exercise:

Problem:

Which of the following methods will provide information to an ecologist about both the size and density of a population?

a. mark and recapture
b. mark and release
c. quadrat
d. life table

Solution:

C

Exercise:

Problem:

Which of the following is best at showing the life expectancy of an individual within a population?

a. quadrat
b. mark and recapture

c. survivorship curve
d. life table

Solution:

D

Exercise:

Problem: Human populations have which type of survivorship curve?

a. Type I
b. Type II
c. Type III
d. Type IV

Solution:

A

Free Response

Exercise:

Problem:

Describe how a researcher would determine the size of a penguin population in Antarctica using the mark and release method.

Solution:

The researcher would mark a certain number of penguins with a tag, release them back into the population, and, at a later time, recapture penguins to see what percentage was tagged. This percentage would allow an estimation of the size of the penguin population.

Glossary

demography
the statistical study of changes in populations over time

life table
a table showing the life expectancy of a population member based on its age

mark and recapture
a method used to determine population size in mobile organisms

mortality rate
the proportion of population surviving to the beginning of an age interval that dies during that age interval

population density
the number of population members divided by the area being measured

population size
the number of individuals in a population

quadrat
a square within which a count of individuals is made that is combined with other such counts to determine population size and density in slow moving or stationary organisms

species distribution pattern
the distribution of individuals within a habitat at a given point in time

survivorship curve
a graph of the number of surviving population members versus the relative age of the member

Population Growth and Regulation
By the end of this section, you will be able to:

- Explain the characteristics of and differences between exponential and logistic growth patterns
- Give examples of exponential and logistic growth in natural populations
- Give examples of how the carrying capacity of a habitat may change
- Compare and contrast density-dependent growth regulation and density-independent growth regulation giving examples

Population ecologists make use of a variety of methods to model population dynamics. An accurate model should be able to describe the changes occurring in a population and predict future changes.

Population Growth

The two simplest models of population growth use deterministic equations (equations that do not account for random events) to describe the rate of change in the size of a population over time. The first of these models, exponential growth, describes theoretical populations that increase in numbers without any limits to their growth. The second model, logistic growth, introduces limits to reproductive growth that become more intense as the population size increases. Neither model adequately describes natural populations, but they provide points of comparison.

Exponential Growth

Charles Darwin, in developing his theory of natural selection, was influenced by the English clergyman Thomas Malthus. Malthus published his book in 1798 stating that populations with abundant natural resources grow very rapidly; however, they limit further growth by depleting their resources. The early pattern of accelerating population size is called **exponential growth**.

The best example of exponential growth in organisms is seen in bacteria. Bacteria are prokaryotes that reproduce largely by binary fission. This division takes about an hour for many bacterial species. If 1000 bacteria are placed in a large flask with an abundant supply of nutrients (so the nutrients will not become quickly depleted), the number of bacteria will have doubled from 1000 to 2000 after just an hour. In another hour, each of the 2000 bacteria will divide, producing 4000 bacteria. After the third hour, there should be 8000 bacteria in the flask. The important concept of exponential growth is that the growth rate—the number of organisms added in each reproductive generation—is itself increasing; that is, the population size is increasing at a greater and greater rate. After 24 of these cycles, the population would have increased from 1000 to more than 16 billion bacteria. When the population size, *N*, is plotted over time, a **J-shaped growth curve** is produced ([link]**a**).

The bacteria-in-a-flask example is not truly representative of the real world where resources are usually limited. However, when a species is introduced into a new habitat that it finds suitable, it may show exponential growth for a while. In the case of the bacteria in the flask, some bacteria will die during the experiment and thus not reproduce; therefore, the growth rate is lowered from a maximal rate in which there is no mortality. The growth rate of a population is largely determined by subtracting the **death rate**, *D*, (number organisms that die during an interval) from the **birth rate**, *B*, (number organisms that are born during an interval). The growth rate can be expressed in a simple equation that combines the birth and death rates into a single factor: *r*. This is shown in the following formula:

Equation:

$$\text{Population growth} = rN$$

The value of *r* can be positive, meaning the population is increasing in size (the rate of change is positive); or negative, meaning the population is decreasing in size; or zero, in which case the population size is unchanging, a condition known as **zero population growth**.

Logistic Growth

Extended exponential growth is possible only when infinite natural resources are available; this is not the case in the real world. Charles Darwin recognized this fact in his description of the "struggle for existence," which states that individuals will compete (with members of their own or other species) for limited resources. The successful ones are more likely to survive and pass on the traits that made them successful to the next generation at a greater rate (natural selection). To model the reality of limited resources, population ecologists developed the **logistic growth** model.

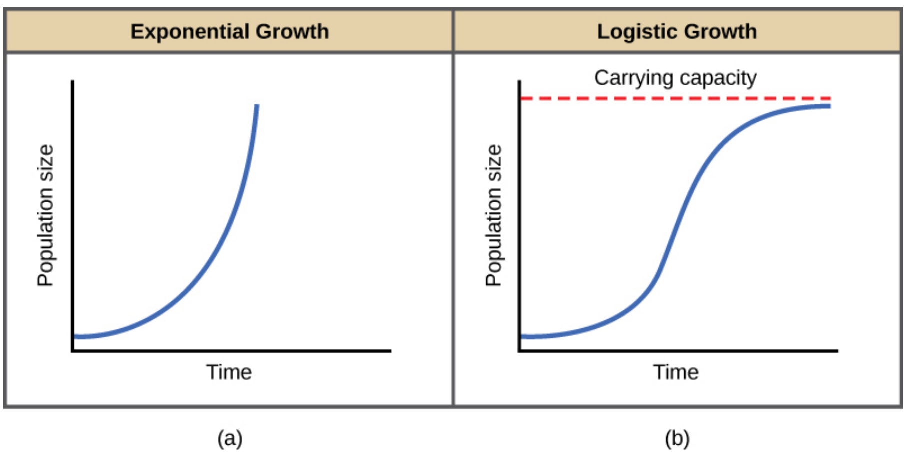
Carrying Capacity and the Logistic Model

In the real world, with its limited resources, exponential growth cannot continue indefinitely. Exponential growth may occur in environments where there are few individuals and plentiful resources, but when the number of individuals gets large enough, resources will be depleted and the growth rate will slow down. Eventually, the growth rate will plateau or level off ([link]**b**). This population size, which is determined by the maximum population size that a particular environment can sustain, is called the **carrying capacity**, or *K*. In real populations, a growing population often overshoots its carrying capacity, and the death rate increases beyond the birth rate causing the population size to decline back to the carrying capacity or below it. Most populations usually fluctuate around the carrying capacity in an undulating fashion rather than existing right at it.

The formula used to calculate logistic growth adds the carrying capacity as a moderating force in the growth rate. The expression "$K – N$" is equal to the number of individuals that may be added to a population at a given time, and "$K – N$" divided by "K" is the fraction of the carrying capacity available for further growth. Thus, the exponential growth model is restricted by this factor to generate the logistic growth equation:
Equation:

$$\text{Population growth} = rN \quad \frac{K - N}{K}$$

Notice that when *N* is almost zero the quantity in brackets is almost equal to 1 (or *K*/*K*) and growth is close to exponential. When the population size is equal to the carrying capacity, or *N* = *K*, the quantity in brackets is equal to zero and growth is equal to zero. A graph of this equation (logistic growth) yields the **S-shaped curve** ([link]**b**). It is a more realistic model of population growth than exponential growth. There are three different sections to an S-shaped curve. Initially, growth is exponential because there are few individuals and ample resources available. Then, as resources begin to become limited, the growth rate decreases. Finally, the growth rate levels off at the carrying capacity of the environment, with little change in population number over time.



When resources are unlimited, populations exhibit (a) exponential growth, shown in a J-shaped curve. When resources are limited, populations exhibit (b) logistic growth. In logistic growth, population expansion decreases as resources become scarce, and it levels off when the carrying capacity of the environment is reached. The logistic growth curve is S-shaped.

Role of Intraspecific Competition

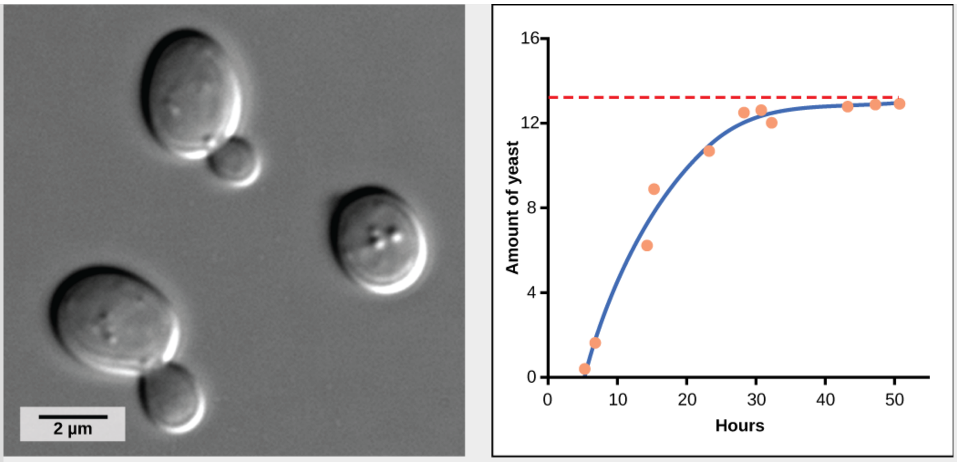
The logistic model assumes that every individual within a population will have equal access to resources and, thus, an equal chance for survival. For plants, the amount of water, sunlight, nutrients, and space to grow are the important resources, whereas in animals, important resources include food, water, shelter, nesting space, and mates.

In the real world, phenotypic variation among individuals within a population means that some individuals will be better adapted to their environment than others. The resulting competition for resources among population members of the same species is termed **intraspecific competition**. Intraspecific competition may not affect populations that are well below their carrying capacity, as resources are plentiful and all individuals can obtain what they need. However, as population size increases, this competition intensifies. In addition, the accumulation of waste products can reduce carrying capacity in an environment.

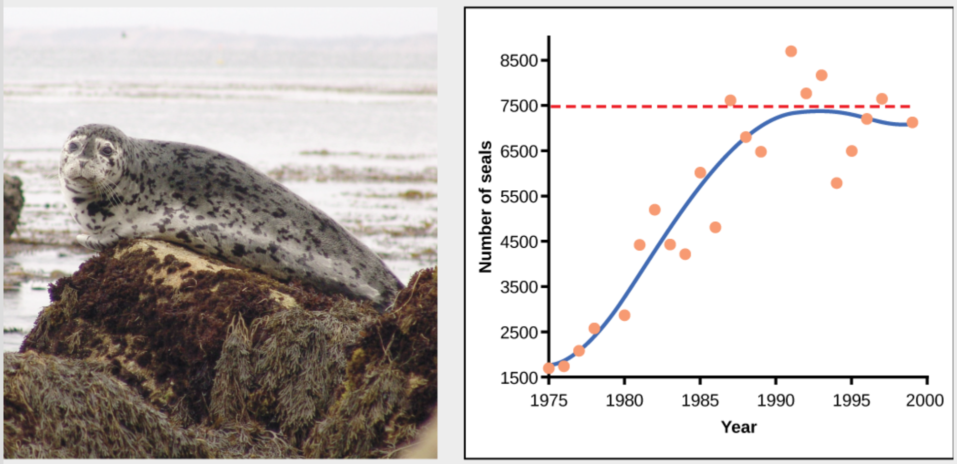
Examples of Logistic Growth

Yeast, a microscopic fungus used to make bread and alcoholic beverages, exhibits the classical S-shaped curve when grown in a test tube ([link]**a**). Its growth levels off as the population depletes the nutrients that are necessary for its growth. In the real world, however, there are variations to this idealized curve. Examples in wild populations include sheep and harbor seals ([link]**b**). In both examples, the population size exceeds the carrying capacity for short periods of time and then falls below the carrying capacity afterwards. This fluctuation in population size continues to occur as the population oscillates around its carrying capacity. Still, even with this oscillation, the logistic model is confirmed.

Note:
Art Connection



(a)



(b)

(a) Yeast grown in ideal conditions in a test tube shows a classical S-shaped logistic growth curve, whereas (b) a natural population of seals shows real-world fluctuation. The yeast is visualized using differential interference contrast light micrography. (credit a: scale-bar data from Matt Russell)

If the major food source of seals declines due to pollution or overfishing, which of the following would likely occur?

a. The carrying capacity of seals would decrease, as would the seal population.

b. The carrying capacity of seals would decrease, but the seal population would remain the same.
c. The number of seal deaths would increase, but the number of births would also increase, so the population size would remain the same.
d. The carrying capacity of seals would remain the same, but the population of seals would decrease.

Population Dynamics and Regulation

The logistic model of population growth, while valid in many natural populations and a useful model, is a simplification of real-world population dynamics. Implicit in the model is that the carrying capacity of the environment does not change, which is not the case. The carrying capacity varies annually. For example, some summers are hot and dry whereas others are cold and wet; in many areas, the carrying capacity during the winter is much lower than it is during the summer. Also, natural events such as earthquakes, volcanoes, and fires can alter an environment and hence its carrying capacity. Additionally, populations do not usually exist in isolation. They share the environment with other species, competing with them for the same resources (interspecific competition). These factors are also important to understanding how a specific population will grow.

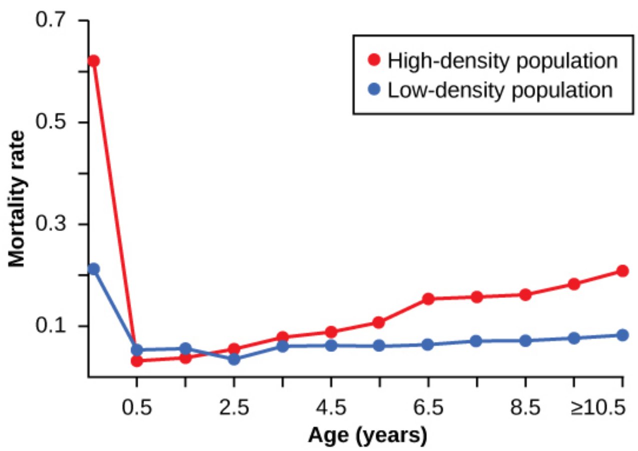
Population growth is regulated in a variety of ways. These are grouped into **density-dependent** factors, in which the density of the population affects growth rate and mortality, and **density-independent** factors, which cause mortality in a population regardless of population density. Wildlife biologists, in particular, want to understand both types because this helps them manage populations and prevent extinction or overpopulation.

Density-dependent Regulation

Most density-dependent factors are biological in nature and include predation, inter- and intraspecific competition, and parasites. Usually, the denser a population is, the greater its mortality rate. For example, during

intra- and interspecific competition, the reproductive rates of the species will usually be lower, reducing their populations' rate of growth. In addition, low prey density increases the mortality of its predator because it has more difficulty locating its food source. Also, when the population is denser, diseases spread more rapidly among the members of the population, which affect the mortality rate.

Density dependent regulation was studied in a natural experiment with wild donkey populations on two sites in Australia.[footnote] On one site the population was reduced by a population control program; the population on the other site received no interference. The high-density plot was twice as dense as the low-density plot. From 1986 to 1987 the high-density plot saw no change in donkey density, while the low-density plot saw an increase in donkey density. The difference in the growth rates of the two populations was caused by mortality, not by a difference in birth rates. The researchers found that numbers of offspring birthed by each mother was unaffected by density. Growth rates in the two populations were different mostly because of juvenile mortality caused by the mother's malnutrition due to scarce high-quality food in the dense population. [link] shows the difference in age-specific mortalities in the two populations.
David Choquenot, "Density-Dependent Growth, Body Condition, and Demography in Feral Donkeys: Testing the Food Hypothesis," *Ecology* 72, no. 3 (June 1991):805–813.



This graph shows the age-specific

mortality rates for wild donkeys from high- and low-density populations. The juvenile mortality is much higher in the high-density population because of maternal malnutrition caused by a shortage of high-quality food.

Density-independent Regulation and Interaction with Density-dependent Factors

Many factors that are typically physical in nature cause mortality of a population regardless of its density. These factors include weather, natural disasters, and pollution. An individual deer will be killed in a forest fire regardless of how many deer happen to be in that area. Its chances of survival are the same whether the population density is high or low. The same holds true for cold winter weather.

In real-life situations, population regulation is very complicated and density-dependent and independent factors can interact. A dense population that suffers mortality from a density-independent cause will be able to recover differently than a sparse population. For example, a population of deer affected by a harsh winter will recover faster if there are more deer remaining to reproduce.

Note:
Evolution in Action
Why Did the Woolly Mammoth Go Extinct?

The three images include: (a) 1916 mural of a mammoth herd from the American Museum of Natural History, (b) the only stuffed mammoth in the world is in the Museum of Zoology located in St. Petersburg, Russia, and (c) a one-month-old baby mammoth, named Lyuba, discovered in Siberia in 2007. (credit a: modification of work by Charles R. Knight; credit b: modification of work by “Tanapon”/Flickr; credit c: modification of work by Matt Howry)

Woolly mammoths began to go extinct about 10,000 years ago, soon after paleontologists believe humans able to hunt them began to colonize North America and northern Eurasia ([link]). A mammoth population survived on Wrangel Island, in the East Siberian Sea, and was isolated from human contact until as recently as 1700 BC. We know a lot about these animals from carcasses found frozen in the ice of Siberia and other northern regions.

It is commonly thought that climate change and human hunting led to their extinction. A 2008 study estimated that climate change reduced the mammoth's range from 3,000,000 square miles 42,000 years ago to 310,000 square miles 6,000 years ago.[footnote] Through archaeological evidence of kill sites, it is also well documented that humans hunted these animals. A 2012 study concluded that no single factor was exclusively responsible for the extinction of these magnificent creatures.[footnote] In addition to climate change and reduction of habitat, scientists demonstrated another important factor in the mammoth's extinction was the migration of human hunters across the Bering Strait to North America during the last ice age 20,000 years ago.

David Nogués-Bravo et al., "Climate Change, Humans, and the Extinction of the Woolly Mammoth." *PLoS Biol* 6 (April 2008): e79, doi:10.1371/journal.pbio.0060079.

G.M. MacDonald et al., "Pattern of Extinction of the Woolly Mammoth in Beringia." *Nature Communications* 3, no. 893 (June 2012), doi:10.1038/ncomms1881.

The maintenance of stable populations was and is very complex, with many interacting factors determining the outcome. It is important to remember that humans are also part of nature. Once we contributed to a species' decline using primitive hunting technology only.

Demographic-Based Population Models

Population ecologists have hypothesized that suites of characteristics may evolve in species that lead to particular adaptations to their environments. These adaptations impact the kind of population growth their species experience. Life history characteristics such as birth rates, age at first reproduction, the numbers of offspring, and even death rates evolve just like anatomy or behavior, leading to adaptations that affect population growth. Population ecologists have described a continuum of life-history "strategies" with *K*-selected species on one end and *r*-selected species on the other. ***K*-selected species** are adapted to stable, predictable environments. Populations of *K*-selected species tend to exist close to their carrying capacity. These species tend to have larger, but fewer, offspring

and contribute large amounts of resources to each offspring. Elephants would be an example of a *K*-selected species. ***r*-selected species** are adapted to unstable and unpredictable environments. They have large numbers of small offspring. Animals that are *r*-selected do not provide a lot of resources or parental care to offspring, and the offspring are relatively self-sufficient at birth. Examples of *r*-selected species are marine invertebrates such as jellyfish and plants such as the dandelion. The two extreme strategies are at two ends of a continuum on which real species life histories will exist. In addition, life history strategies do not need to evolve as suites, but can evolve independently of each other, so each species may have some characteristics that trend toward one extreme or the other.

Section Summary

Populations with unlimited resources grow exponentially—with an accelerating growth rate. When resources become limiting, populations follow a logistic growth curve in which population size will level off at the carrying capacity.

Populations are regulated by a variety of density-dependent and density-independent factors. Life-history characteristics, such as age at first reproduction or numbers of offspring, are characteristics that evolve in populations just as anatomy or behavior can evolve over time. The model of *r*- and *K*-selection suggests that characters, and possibly suites of characters, may evolve adaptations to population stability near the carrying capacity (*K*-selection) or rapid population growth and collapse (*r*-selection). Species will exhibit adaptations somewhere on a continuum between these two extremes.

Art Exercise

Exercise:

Problem:

[link] If the major food source of seals declines due to pollution or overfishing, which of the following would likely occur?

a. The carrying capacity of seals would decrease, as would the seal population.
b. The carrying capacity of seals would decrease, but the seal population would remain the same.
c. The number of seal deaths would increase, but the number of births would also increase, so the population size would remain the same.
d. The carrying capacity of seals would remain the same, but the population of seals would decrease.

Solution:

[link] A: The carrying capacity of seals would decrease, as would the seal population.

Multiple Choice

Exercise:

Problem:

Species with limited resources usually exhibit a(n) _________ growth curve.

a. logistic
b. logical
c. experimental
d. exponential

Solution:

A

Exercise:

Problem:

The maximum growth rate characteristic of a species is called its _________.

a. limit
b. carrying capacity
c. biotic potential
d. exponential growth pattern

Solution:

C

Exercise:

Problem:

The population size of a species capable of being supported by the environment is called its _________.

a. limit
b. carrying capacity
c. biotic potential
d. logistic growth pattern

Solution:

B

Exercise:

Problem: Species that have many offspring at one time are usually:

a. *r*-selected
b. *K*-selected
c. both *r*- and *K*-selected

d. not selected

Solution:

A

Exercise:

Problem: A forest fire is an example of _________ regulation.

a. density-dependent
b. density-independent
c. *r*-selected
d. *K*-selected

Solution:

B

Free Response

Exercise:

Problem:

Describe the growth at various parts of the S-shaped curve of logistic growth.

Solution:

In the first part of the curve, when few individuals of the species are present and resources are plentiful, growth is exponential, similar to a J-shaped curve. Later, growth slows due to the species using up resources. Finally, the population levels off at the carrying capacity of the environment, and it is relatively stable over time.

Exercise:

Problem:

Give an example of how density-dependent and density-independent factors might interact.

Solution:

If a natural disaster such as a fire happened in the winter, when populations are low, it would have a greater effect on the overall population and its recovery than if the same disaster occurred during the summer, when population levels are high.

Glossary

birth rate
: the number of births within a population at a specific point in time

carrying capacity
: the maximum number of individuals of a population that can be supported by the limited resources of a habitat

death rate
: the number of deaths within a population at a specific point in time

density-dependent regulation
: the regulation of population in which birth and death rates are dependent on population size

density-independent regulation
: the regulation of population in which the death rate is independent of the population size

exponential growth
: an accelerating growth pattern seen in populations where resources are not limiting

intraspecific competition

the competition among members of the same species

J-shaped growth curve
the shape of an exponential growth curve

K-selected species
a species suited to stable environments that produce a few, relatively large offspring and provide parental care

logistic growth
the leveling off of exponential growth due to limiting resources

r-selected species
a species suited to changing environments that produce many offspring and provide little or no parental care

S-shaped growth curve
the shape of a logistic growth curve

zero population growth
the steady population size where birth rates and death rates are equal

The Human Population

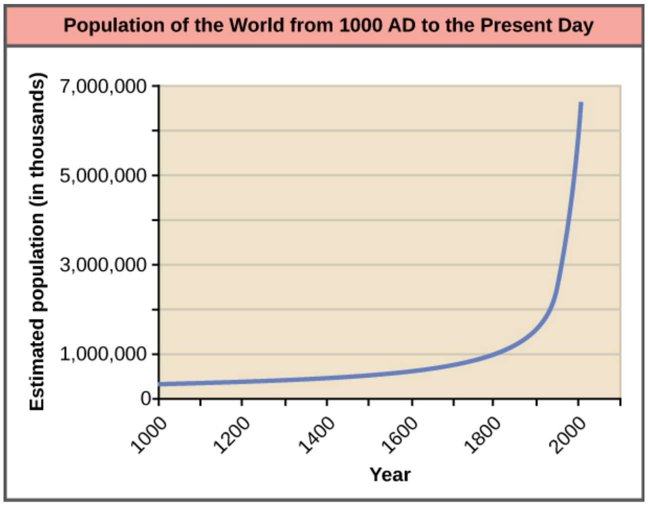
By the end of this section, you will be able to:

- Discuss how human population growth can be exponential
- Explain how humans have expanded the carrying capacity of their habitat
- Relate population growth and age structure to the level of economic development in different countries
- Discuss the long-term implications of unchecked human population growth

Concepts of animal population dynamics can be applied to human population growth. Humans are not unique in their ability to alter their environment. For example, beaver dams alter the stream environment where they are built. Humans, however, have the ability to alter their environment to increase its carrying capacity, sometimes to the detriment of other species. Earth’s human population and their use of resources are growing rapidly, to the extent that some worry about the ability of Earth’s environment to sustain its human population. Long-term exponential growth carries with it the potential risks of famine, disease, and large-scale death, as well as social consequences of crowding such as increased crime.

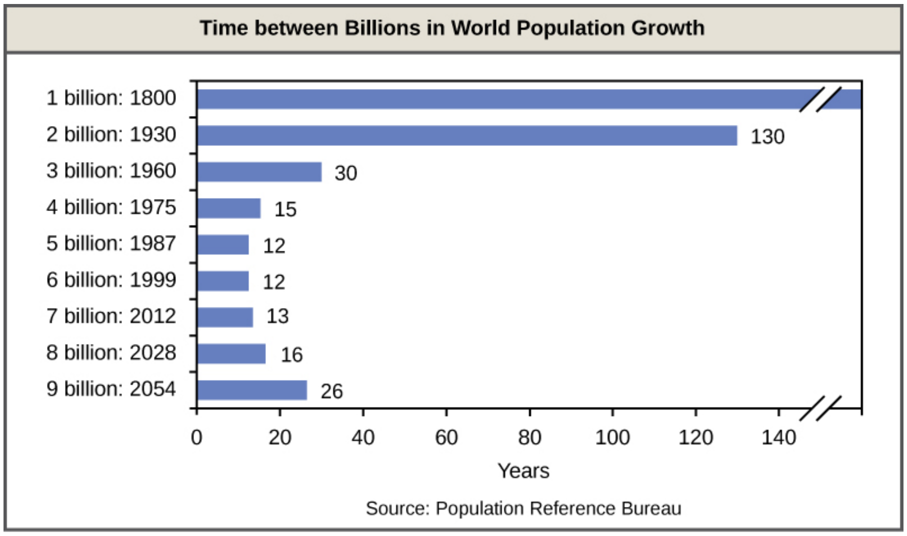
Human technology and particularly our harnessing of the energy contained in fossil fuels have caused unprecedented changes to Earth’s environment, altering ecosystems to the point where some may be in danger of collapse. Changes on a global scale including depletion of the ozone layer, desertification and topsoil loss, and global climate change are caused by human activities.

The world’s human population is presently growing exponentially ([link]).



Human population growth since 1000 AD is exponential.

A consequence of exponential growth rate is that the time that it takes to add a particular number of humans to the population is becoming shorter. [link] shows that 123 years were necessary to add 1 billion humans between 1804 and 1930, but it only took 24 years to add the two billion people between 1975 and 1999. This acceleration in growth rate will likely begin to decrease in the coming decades. Despite this, the population will continue to increase and the threat of overpopulation remains, particularly because the damage caused to ecosystems and biodiversity is lowering the human carrying capacity of the planet.



The time between the addition of each billion human beings to Earth decreases over time. (credit: modification of work by Ryan T. Cragun)

Note:
Concept in Action

Click through this interactive view of how human populations have changed over time.

Overcoming Density-Dependent Regulation

Humans are unique in their ability to alter their environment in myriad ways. This ability is responsible for human population growth because it resets the carrying capacity and overcomes density-dependent growth regulation. Much of this ability is related to human intelligence, society, and communication. Humans construct shelters to protect themselves from the elements and have developed agriculture and domesticated animals to increase their food supplies. In addition, humans use language to communicate this technology to new generations, allowing them to improve upon previous accomplishments.

Other factors in human population growth are migration and public health. Humans originated in Africa, but we have since migrated to nearly all inhabitable land on Earth, thus, increasing the area that we have colonized. Public health, sanitation, and the use of antibiotics and vaccines have decreased the ability of infectious disease to limit human population growth in developed countries. In the past, diseases such as the bubonic plaque of the fourteenth century killed between 30 and 60 percent of Europe's population and reduced the overall world population by as many as one hundred million people. Infectious disease continues to have an impact on human population growth. For example, life expectancy in sub-Saharan Africa, which was increasing from 1950 to 1990, began to decline after 1985 largely as a result of HIV/AIDS mortality. The reduction in life expectancy caused by HIV/AIDS was estimated to be 7 years for 2005. [footnote]

Danny Dorling, Mary Shaw, and George Davey Smith, "Global Inequality of Life Expectancy due to AIDS," *BMJ* 332, no. 7542 (March 2006): 662-664, doi: 10.1136/bmj.332.7542.662.

Declining life expectancy is an indicator of higher mortality rates and leads to lower birth rates.

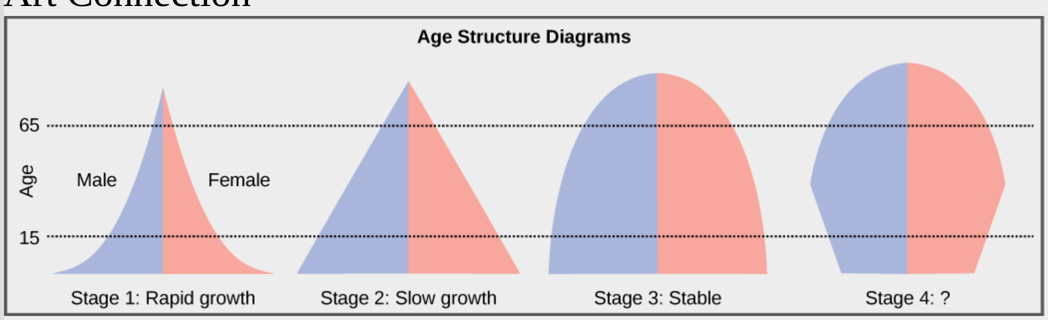
The fundamental cause of the acceleration of growth rate for humans in the past 200 years has been the reduced death rate due to a development of the technological advances of the industrial age, urbanization that supported those technologies, and especially the exploitation of the energy in fossil fuels. Fossil fuels are responsible for dramatically increasing the resources

available for human population growth through agriculture (mechanization, pesticides, and fertilizers) and harvesting wild populations.

Age Structure, Population Growth, and Economic Development

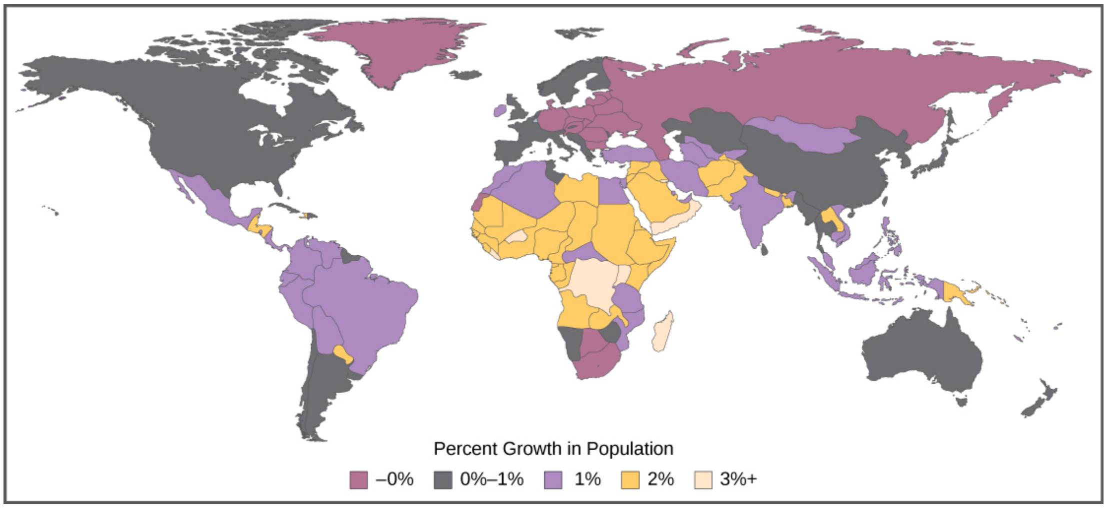
The age structure of a population is an important factor in population dynamics. **Age structure** is the proportion of a population in different age classes. Models that incorporate age structure allow better prediction of population growth, plus the ability to associate this growth with the level of economic development in a region. Countries with rapid growth have a pyramidal shape in their age structure diagrams, showing a preponderance of younger individuals, many of whom are of reproductive age ([link]). This pattern is most often observed in underdeveloped countries where individuals do not live to old age because of less-than-optimal living conditions, and there is a high birth rate. Age structures of areas with slow growth, including developed countries such as the United States, still have a pyramidal structure, but with many fewer young and reproductive-aged individuals and a greater proportion of older individuals. Other developed countries, such as Italy, have zero population growth. The age structure of these populations is more conical, with an even greater percentage of middle-aged and older individuals. The actual growth rates in different countries are shown in [link], with the highest rates tending to be in the less economically developed countries of Africa and Asia.

Note:
Art Connection



Typical age structure diagrams are shown. The rapid growth diagram narrows to a point, indicating that the number of individuals decreases rapidly with age. In the slow growth model, the number of individuals decreases steadily with age. Stable population diagrams are rounded on the top, showing that the number of individuals per age group decreases gradually, and then increases for the older part of the population.

Age structure diagrams for rapidly growing, slow growing, and stable populations are shown in stages 1 through 3. What type of population change do you think stage 4 represents?



The percent growth rate of population in different countries is shown. Notice that the highest growth is occurring in less economically developed countries in Africa and Asia.

Long-Term Consequences of Exponential Human Population Growth

Many dire predictions have been made about the world's population leading to a major crisis called the "population explosion." In the 1968 book *The Population Bomb*, biologist Dr. Paul R. Ehrlich wrote, "The battle to feed all of humanity is over. In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date nothing can prevent a substantial increase in the world death rate."[footnote] While many critics view this statement as an exaggeration, the laws of exponential population growth are still in effect, and unchecked human population growth cannot continue indefinitely.

Paul R. Erlich, prologue to *The Population Bomb*, (1968; repr., New York: Ballantine, 1970).

Efforts to moderate population control led to the **one-child policy** in China, which imposes fines on urban couples who have more than one child. Due to the fact that some couples wish to have a male heir, many Chinese couples continue to have more than one child. The effectiveness of the policy in limiting overall population growth is controversial, as is the policy itself. Moreover, there are stories of female infanticide having occurred in some of the more rural areas of the country. Family planning education programs in other countries have had highly positive effects on limiting population growth rates and increasing standards of living. In spite of population control policies, the human population continues to grow. Because of the subsequent need to produce more and more food to feed our population, inequalities in access to food and other resources will continue to widen. The United Nations estimates the future world population size could vary from 6 billion (a decrease) to 16 billion people by the year 2100. There is no way to know whether human population growth will moderate to the point where the crisis described by Dr. Ehrlich will be averted.

Another consequence of population growth is the change and degradation of the natural environment. Many countries have attempted to reduce the human impact on climate change by limiting their emission of greenhouse gases. However, a global climate change treaty remains elusive, and many underdeveloped countries trying to improve their economic condition may be less likely to agree with such provisions without compensation if it means slowing their economic development. Furthermore, the role of human activity in causing climate change has become a hotly debated

socio-political issue in some developed countries, including the United States. Thus, we enter the future with considerable uncertainty about our ability to curb human population growth and protect our environment to maintain the carrying capacity for the human species.

Note:
Concept in Action

Visit this website and select “Launch the movie” for an animation discussing the global impacts of human population growth.

Section Summary

Earth’s human population is growing exponentially. Humans have increased their carrying capacity through technology, urbanization, and harnessing the energy of fossil fuels. The age structure of a population allows us to predict population growth. Unchecked human population growth could have dire long-term effects on human welfare and Earth’s ecosystems.

Art Connections

Exercise:

Problem:

[link] Age structure diagrams for rapidly growing, slow growing, and stable populations are shown in stages 1 through 3. What type of population change do you think stage 4 represents?

Solution:

[link] Stage 4 represents a population that is decreasing.

Multiple Choice

Exercise:

Problem:

A country with zero population growth is likely to be ________.

a. in Africa
b. in Asia
c. economically developed
d. economically underdeveloped

Solution:

C

Exercise:

Problem:

Which type of country has the greatest proportion of young individuals?

a. economically developed
b. economically underdeveloped
c. countries with zero population growth
d. countries in Europe

Solution:

B

Exercise:

Problem:

Which of the following is not a way that humans have increased the carrying capacity of the environment?

a. agriculture
b. using large amounts of natural resources
c. domestication of animals
d. use of language

Solution:

B

Free Response

Exercise:

Problem:

Describe the age structures in rapidly growing countries, slowly growing countries, and countries with zero population growth.

Solution:

Rapidly growing countries have a large segment of the population at reproductive age or younger. Slower growing populations have a lower percentage of these individuals, and countries with zero population growth have an even lower percentage. On the other hand, a high proportion of older individuals is seen mostly in countries with zero growth, and a low proportion is most common in rapidly growing countries.

Glossary

age structure

the distribution of the proportion of population members in each age class

one-child policy

a policy in China to limit population growth by limiting urban couples to have only one child or face a penalty of a fine

Community Ecology
By the end of this section, you will be able to:

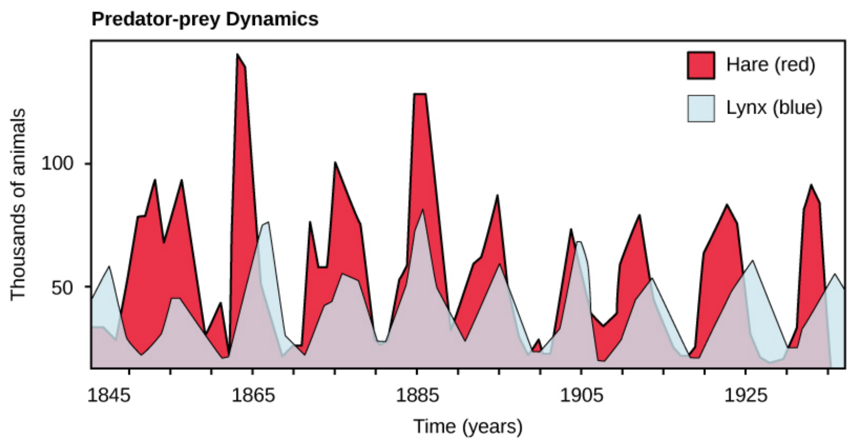
- Discuss the predator-prey cycle
- Give examples of defenses against predation and herbivory
- Describe the competitive exclusion principle
- Give examples of symbiotic relationships between species
- Describe community structure and succession

In general, populations of one species never live in isolation from populations of other species. The interacting populations occupying a given habitat form an ecological community. The number of species occupying the same habitat and their relative abundance is known as the diversity of the community. Areas with low species diversity, such as the glaciers of Antarctica, still contain a wide variety of living organisms, whereas the diversity of tropical rainforests is so great that it cannot be accurately assessed. Scientists study ecology at the community level to understand how species interact with each other and compete for the same resources.

Predation and Herbivory

Perhaps the classical example of species interaction is the predator-prey relationship. The narrowest definition of the predator-prey interaction describes individuals of one population that kill and then consume the individuals of another population. Population sizes of predators and prey in a community are not constant over time, and they may vary in cycles that appear to be related. The most often cited example of predator-prey population dynamics is seen in the cycling of the lynx (predator) and the snowshoe hare (prey), using 100 years of trapping data from North America ([link]). This cycling of predator and prey population sizes has a period of approximately ten years, with the predator population lagging one to two years behind the prey population. An apparent explanation for this pattern is that as the hare numbers increase, there is more food available for the lynx, allowing the lynx population to increase as well. When the lynx population grows to a threshold level, however, they kill so many hares that hare numbers begin to decline, followed by a decline in the lynx population because of scarcity of food. When the lynx population is low, the hare

population size begins to increase due, in part, to low predation pressure, starting the cycle anew.



The cycling of snowshoe hare and lynx populations in Northern Ontario is an example of predator-prey dynamics.

Defense Mechanisms against Predation and Herbivory

Predation and predator avoidance are strong selective agents. Any heritable character that allows an individual of a prey population to better evade its predators will be represented in greater numbers in later generations. Likewise, traits that allow a predator to more efficiently locate and capture its prey will lead to a greater number of offspring and an increase in the commonness of the trait within the population. Such ecological relationships between specific populations lead to adaptations that are driven by reciprocal evolutionary responses in those populations. Species have evolved numerous mechanisms to escape predation and herbivory (the consumption of plants for food). Defenses may be mechanical, chemical, physical, or behavioral.

Mechanical defenses, such as the presence of armor in animals or thorns in plants, discourage predation and herbivory by discouraging physical contact ([link]**a**). Many animals produce or obtain chemical defenses from plants and store them to prevent predation. Many plant species produce secondary plant compounds that serve no function for the plant except that they are toxic to animals and discourage consumption. For example, the foxglove produces several compounds, including digitalis, that are extremely toxic when eaten ([link]**b**). (Biomedical scientists have purposed the chemical produced by foxglove as a heart medication, which has saved lives for many decades.)

(a) (b)

The (a) honey locust tree uses thorns, a mechanical defense, against herbivores, while the (b) foxglove uses a chemical defense: toxins produces by the plant can cause nausea, vomiting, hallucinations, convulsions, or death when consumed. (credit a: modification of work by Huw Williams; credit b: modification of work by Philip Jägenstedt)

Many species use their body shape and coloration to avoid being detected by predators. The tropical walking stick is an insect with the coloration and body shape of a twig, which makes it very hard to see when it is stationary

against a background of real twigs ([link]**a**). In another example, the chameleon can change its color to match its surroundings ([link]**b**).

(a) The tropical walking stick and (b) the chameleon use their body shape and/or coloration to prevent detection by predators. (credit a: modification of work by Linda Tanner; credit b: modification of work by Frank Vassen)

Some species use coloration as a way of warning predators that they are distasteful or poisonous. For example, the monarch butterfly caterpillar sequesters poisons from its food (plants and milkweeds) to make itself poisonous or distasteful to potential predators. The caterpillar is bright yellow and black to advertise its toxicity. The caterpillar is also able to pass the sequestered toxins on to the adult monarch, which is also dramatically colored black and red as a warning to potential predators. Fire-bellied toads produce toxins that make them distasteful to their potential predators. They have bright red or orange coloration on their bellies, which they display to a potential predator to advertise their poisonous nature and discourage an attack. These are only two examples of warning coloration, which is a relatively common adaptation. Warning coloration only works if a predator uses eyesight to locate prey and can learn—a naïve predator must experience the negative consequences of eating one before it will avoid other similarly colored individuals ([link]).

The fire-bellied toad has bright coloration on its belly that serves to warn potential predators that it is toxic. (credit: modification of work by Roberto Verzo)

While some predators learn to avoid eating certain potential prey because of their coloration, other species have evolved mechanisms to mimic this coloration to avoid being eaten, even though they themselves may not be unpleasant to eat or contain toxic chemicals. In some cases of **mimicry**, a harmless species imitates the warning coloration of a harmful species. Assuming they share the same predators, this coloration then protects the harmless ones. Many insect species mimic the coloration of wasps, which are stinging, venomous insects, thereby discouraging predation ([link]).

(a) (b)

One form of mimicry is when a harmless species mimics the coloration of a harmful species, as is seen with the (a) wasp (*Polistes* sp.) and the (b) hoverfly (*Syrphus* sp.). (credit: modification of work by Tom Ings)

In other cases of mimicry, multiple species share the same warning coloration, but all of them actually have defenses. The commonness of the signal improves the compliance of all the potential predators. [link] shows a variety of foul-tasting butterflies with similar coloration.

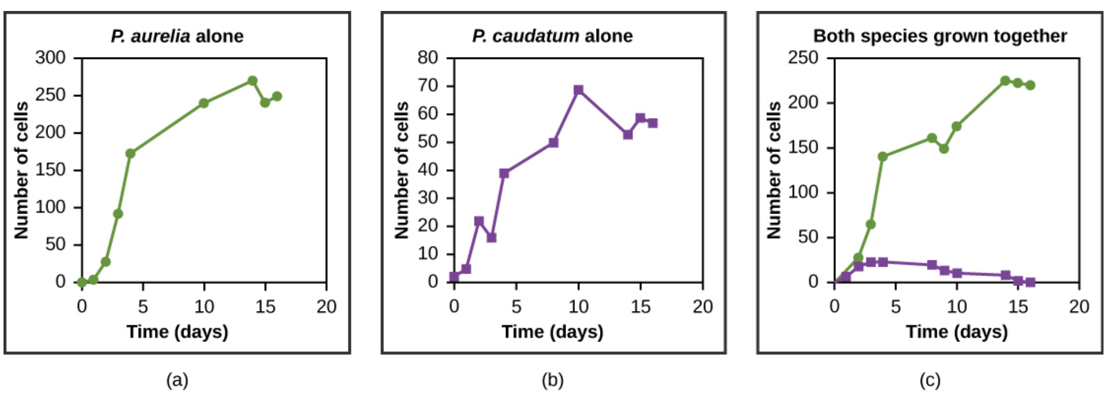
Several unpleasant-tasting *Heliconius* butterfly species share a similar color pattern with better-tasting varieties, an example of mimicry. (credit: Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, et al.)

Note:
Concept in Action

Go to this website to view stunning examples of mimicry.

Competitive Exclusion Principle

Resources are often limited within a habitat and multiple species may compete to obtain them. Ecologists have come to understand that all species have an ecological niche. A niche is the unique set of resources used by a species, which includes its interactions with other species. The **competitive exclusion principle** states that two species cannot occupy the same niche in a habitat: in other words, different species cannot coexist in a community if they are competing for all the same resources. This principle works because if there is an overlap in resource use and therefore competition between two species, then traits that lessen reliance on the shared resource will be selected for leading to evolution that reduces the overlap. If either species is unable to evolve to reduce competition, then the species that most efficiently exploits the resource will drive the other species to extinction. An experimental example of this principle is shown in [link] with two protozoan species: *Paramecium aurelia* and *Paramecium caudatum*. When grown individually in the laboratory, they both thrive. But when they are placed together in the same test tube (habitat), *P. aurelia* outcompetes *P. caudatum* for food, leading to the latter's eventual extinction.



Paramecium aurelia and *Paramecium caudatum* grow well individually, but when they compete for the same resources, the *P. aurelia* outcompetes the *P. caudatum*.

Symbiosis

Symbiotic relationships are close, long-term interactions between individuals of different species. Symbioses may be commensal, in which one species benefits while the other is neither harmed nor benefited; mutualistic, in which both species benefit; or parasitic, in which the interaction harms one species and benefits the other.

Commensalism

A commensal relationship occurs when one species benefits from a close prolonged interaction, while the other neither benefits nor is harmed. Birds nesting in trees provide an example of a commensal relationship ([link]). The tree is not harmed by the presence of the nest among its branches. The nests are light and produce little strain on the structural integrity of the branch, and most of the leaves, which the tree uses to get energy by photosynthesis, are above the nest so they are unaffected. The bird, on the other hand, benefits greatly. If the bird had to nest in the open, its eggs and young would be vulnerable to predators. Many potential commensal relationships are difficult to identify because it is difficult to prove that one partner does not derive some benefit from the presence of the other.

The southern masked-weaver is starting to make a nest in a tree in Zambezi Valley, Zambia. This is an example of a commensal relationship, in which one species (the bird) benefits, while the other (the tree) neither benefits nor is harmed. (credit: “Hanay”/Wikimedia Commons)

Mutualism

A second type of symbiotic relationship is called **mutualism**, in which two species benefit from their interaction. For example, termites have a mutualistic relationship with protists that live in the insect’s gut ([link]**a**). The termite benefits from the ability of the protists to digest cellulose. However, the protists are able to digest cellulose only because of the presence of symbiotic bacteria within their cells that produce the cellulase enzyme. The termite itself cannot do this: without the protozoa, it would not be able to obtain energy from its food (cellulose from the wood it chews and eats). The protozoa benefit by having a protective environment and a constant supply of food from the wood chewing actions of the termite. In turn, the protists benefit from the enzymes provided by their bacterial endosymbionts, while the bacteria benefit from a doubly protective environment and a constant source of nutrients from two hosts. Lichen are a mutualistic relationship between a fungus and photosynthetic algae or cyanobacteria ([link]**b**). The glucose produced by the algae provides nourishment for both organisms, whereas the physical structure of the lichen protects the algae from the elements and makes certain nutrients in the atmosphere more available to the algae. The algae of lichens can live

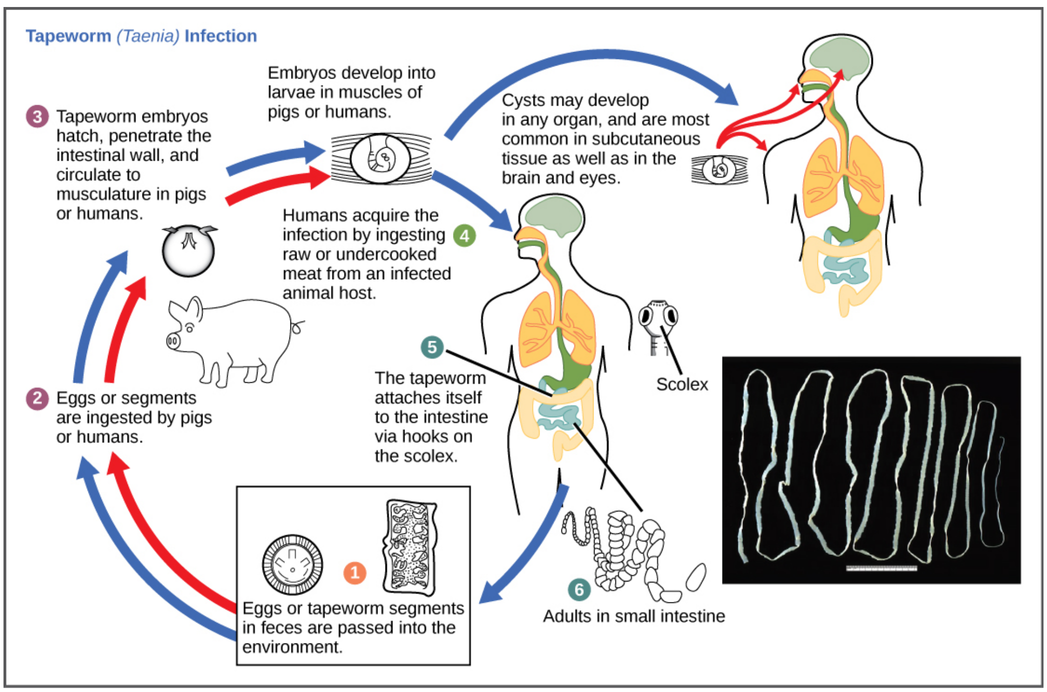
independently given the right environment, but many of the fungal partners are unable to live on their own.

(a) Termites form a mutualistic relationship with symbiotic protozoa in their guts, which allow both organisms to obtain energy from the cellulose the termite consumes. (b) Lichen is a fungus that has symbiotic photosynthetic algae living in close association. (credit a: modification of work by Scott Bauer, USDA; credit b: modification of work by Cory Zanker)

Parasitism

A **parasite** is an organism that feeds off another without immediately killing the organism it is feeding on. In this relationship, the parasite benefits, but the organism being fed upon, the **host**, is harmed. The host is usually weakened by the parasite as it siphons resources the host would normally use to maintain itself. Parasites may kill their hosts, but there is usually selection to slow down this process to allow the parasite time to complete its reproductive cycle before it or its offspring are able to spread to another host.

The reproductive cycles of parasites are often very complex, sometimes requiring more than one host species. A tapeworm causes disease in humans when contaminated, undercooked meat such as pork, fish, or beef is consumed ([link]). The tapeworm can live inside the intestine of the host for several years, benefiting from the host's food, and it may grow to be over 50 feet long by adding segments. The parasite moves from one host species to a second host species in order to complete its life cycle. *Plasmodium falciparum* is another parasite: the protists that cause malaria, a significant disease in many parts of the world. Living inside human liver and red blood cells, the organism reproduces asexually in the human host and then sexually in the gut of blood-feeding mosquitoes to complete its life cycle. Thus malaria is spread from human to mosquito and back to human, one of many arthropod-borne infectious diseases of humans.



This diagram shows the life cycle of the tapeworm, a human worm parasite. (credit: modification of work by CDC)

Note:
Concept in Action

To learn more about “Symbiosis in the Sea,” watch this webisode of Jonathan Bird’s Blue World.

Characteristics of Communities

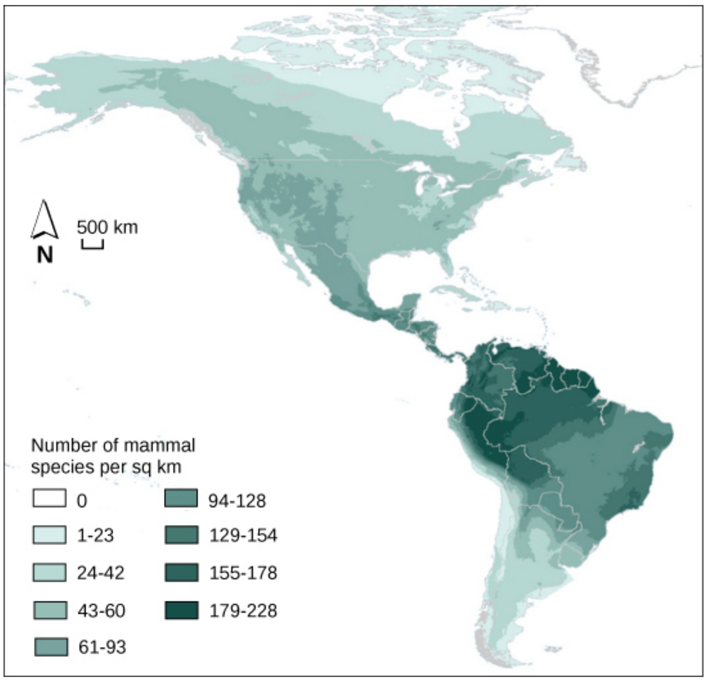
Communities are complex systems that can be characterized by their structure (the number and size of populations and their interactions) and dynamics (how the members and their interactions change over time). Understanding community structure and dynamics allows us to minimize impacts on ecosystems and manage ecological communities we benefit from.

Biodiversity

Ecologists have extensively studied one of the fundamental characteristics of communities: biodiversity. One measure of biodiversity used by ecologists is the number of different species in a particular area and their relative abundance. The area in question could be a habitat, a biome, or the entire biosphere. **Species richness** is the term used to describe the number of species living in a habitat or other unit. Species richness varies across the globe ([link]). Ecologists have struggled to understand the determinants of biodiversity. Species richness is related to latitude: the greatest species richness occurs near the equator and the lowest richness occurs near the poles. Other factors influence species richness as well. **Island biogeography** attempts to explain the great species richness found in

isolated islands, and has found relationships between species richness, island size, and distance from the mainland.

Relative species abundance is the number individuals in a species relative to the total number of individuals in all species within a system. Foundation species, described below, often have the highest relative abundance of species.

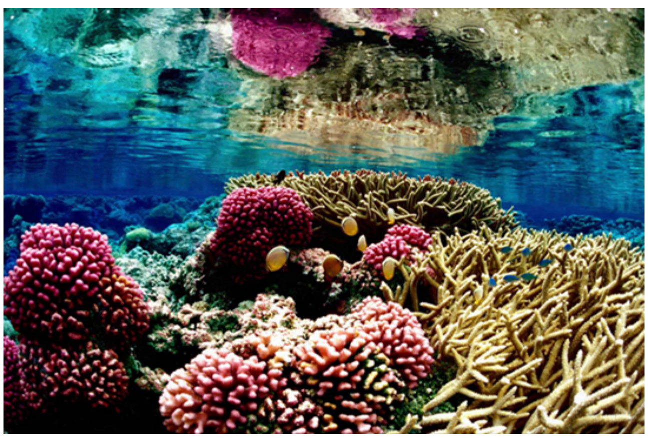


The greatest species richness for mammals in North America is associated in the equatorial latitudes. (credit: modification of work by NASA, CIESIN, Columbia University)

Foundation Species

Foundation species are considered the “base” or “bedrock” of a community, having the greatest influence on its overall structure. They are often primary producers, and they are typically an abundant organism. For example, kelp, a species of brown algae, is a foundation species that forms the basis of the kelp forests off the coast of California.

Foundation species may physically modify the environment to produce and maintain habitats that benefit the other organisms that use them. Examples include the kelp described above or tree species found in a forest. The photosynthetic corals of the coral reef also provide structure by physically modifying the environment ([link]). The exoskeletons of living and dead coral make up most of the reef structure, which protects many other species from waves and ocean currents.



Coral is the foundation species of coral reef ecosystems. (credit: Jim E. Maragos, USFWS)

Keystone Species

A **keystone species** is one whose presence has inordinate influence in maintaining the prevalence of various species in an ecosystem, the

ecological community’s structure, and sometimes its biodiversity. *Pisaster ochraceus*, the intertidal sea star, is a keystone species in the northwestern portion of the United States ([link]). Studies have shown that when this organism is removed from communities, mussel populations (their natural prey) increase, which completely alters the species composition and reduces biodiversity. Another keystone species is the banded tetra, a fish in tropical streams, which supplies nearly all of the phosphorus, a necessary inorganic nutrient, to the rest of the community. The banded tetra feeds largely on insects from the terrestrial ecosystem and then excretes phosphorus into the aquatic ecosystem. The relationships between populations in the community, and possibly the biodiversity, would change dramatically if these fish were to become extinct.

The *Pisaster ochraceus* sea star is a keystone species. (credit: Jerry Kirkhart)

Note:
Biology in Action
Invasive Species

Invasive species are non-native organisms that, when introduced to an area out of its native range, alter the community they invade. In the United States, invasive species like the purple loosestrife (*Lythrum salicaria*) and the zebra mussel (*Dreissena polymorpha*) have altered aquatic ecosystems, and some forests are threatened by the spread of common buckthorn (*Rhamnus cathartica*) and garlic mustard (*Alliaria petiolata*). Some well-known invasive animals include the emerald ash borer (*Agrilus planipennis*) and the European starling (*Sturnus vulgaris*). Whether enjoying a forest hike, taking a summer boat trip, or simply walking down an urban street, you have likely encountered an invasive species.

One of the many recent proliferations of an invasive species concerns the Asian carp in the United States. Asian carp were introduced to the United States in the 1970s by fisheries (commercial catfish ponds) and by sewage treatment facilities that used the fish’s excellent filter feeding abilities to clean their ponds of excess plankton. Some of the fish escaped, and by the 1980s they had colonized many waterways of the Mississippi River basin, including the Illinois and Missouri Rivers.

Voracious feeders and rapid reproducers, Asian carp may outcompete native species for food and could lead to their extinction. One species, the grass carp, feeds on phytoplankton and aquatic plants. It competes with native species for these resources and alters nursery habitats for other fish by removing aquatic plants. Another species, the silver carp, competes with native fish that feed on zooplankton. In some parts of the Illinois River, Asian carp constitute 95 percent of the community's biomass. Although edible, the fish is bony and not desired in the United States. Moreover, their presence now threatens the native fish and fisheries of the Great Lakes, which are important to local economies and recreational anglers. Asian carp have even injured humans. The fish, frightened by the sound of approaching motorboats, thrust themselves into the air, often landing in the boat or directly hitting boaters.

The Great Lakes and their prized salmon and lake trout fisheries are being threatened by Asian carp. The carp are not yet present in the Great Lakes, and attempts are being made to prevent its access to the lakes through the Chicago Ship and Sanitary Canal, which is the only connection between the Mississippi River and Great Lakes basins. To prevent the Asian carp from leaving the canal, a series of electric barriers have been used to discourage their migration; however, the threat is significant enough that

several states and Canada have sued to have the Chicago channel permanently cut off from Lake Michigan. Local and national politicians have weighed in on how to solve the problem. In general, governments have been ineffective in preventing or slowing the introduction of invasive species.

The issues associated with Asian carp show how population and community ecology, fisheries management, and politics intersect on issues of vital importance to the human food supply and economy. Socio-political issues like the Asian carp make extensive use of the sciences of population ecology, the study of members of a particular species occupying a habitat; and community ecology, the study of the interaction of all species within a habitat.

Community Dynamics

Community dynamics are the changes in community structure and composition over time, often following **environmental disturbances** such as volcanoes, earthquakes, storms, fires, and climate change. Communities with a relatively constant number of species are said to be at equilibrium. The equilibrium is dynamic with species identities and relationships changing over time, but maintaining relatively constant numbers. Following a disturbance, the community may or may not return to the equilibrium state.

Succession describes the sequential appearance and disappearance of species in a community over time after a severe disturbance. In **primary succession**, newly exposed or newly formed rock is colonized by living organisms; in **secondary succession**, a part of an ecosystem is disturbed and remnants of the previous community remain. In both cases, there is a sequential change in species until a more or less permanent community develops.

Primary Succession and Pioneer Species

Primary succession occurs when new land is formed, for example, following the eruption of volcanoes, such as those on the Big Island of Hawaii. As lava flows into the ocean, new land is continually being formed. On the Big Island, approximately 32 acres of land is added to it its size each year. Weathering and other natural forces break down the rock enough for the establishment of hearty species such as lichens and some plants, known as **pioneer species** ([link]). These species help to further break down the mineral-rich lava into soil where other, less hardy but more competitive species, such as grasses, shrubs, and trees, will grow and eventually replace the pioneer species. Over time the area will reach an equilibrium state, with a set of organisms quite different from the pioneer species.

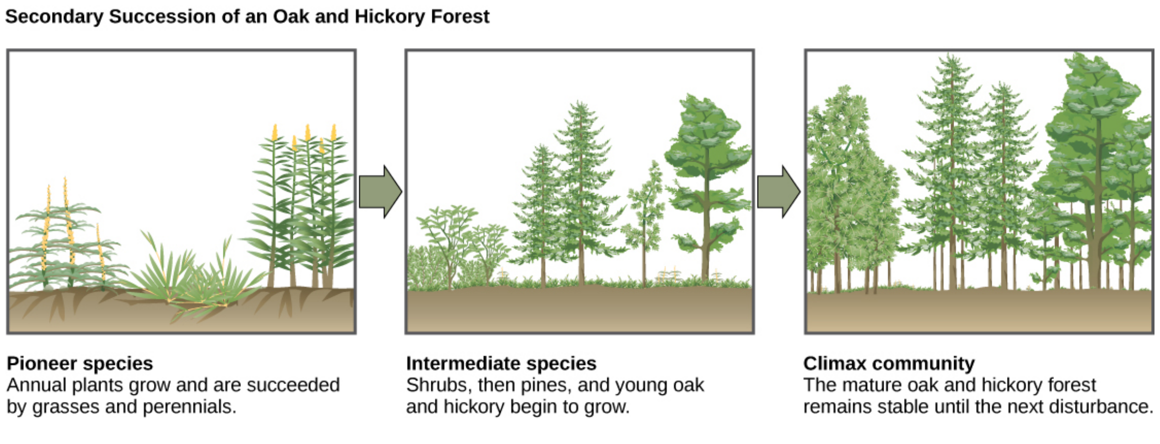
During primary succession in lava on Maui, Hawaii, succulent plants are the pioneer species. (credit: Forest and Kim Starr)

Secondary succession

A classic example of secondary succession occurs in oak and hickory forests cleared by wildfire ([link]). Wildfires will burn most vegetation, and

unless the animals can flee the area, they are killed. Their nutrients, however, are returned to the ground in the form of ash. Thus, although the community has been dramatically altered, there is a soil ecosystem present that provides a foundation for rapid recolonization.

Before the fire, the vegetation was dominated by tall trees with access to the major plant energy resource: sunlight. Their height gave them access to sunlight while also shading the ground and other low-lying species. After the fire, though, these trees are no longer dominant. Thus, the first plants to grow back are usually annual plants followed within a few years by quickly growing and spreading grasses and other pioneer species. Due, at least in part, to changes in the environment brought on by the growth of grasses and forbs, over many years, shrubs emerge along with small pine, oak, and hickory trees. These organisms are called intermediate species. Eventually, over 150 years, the forest will reach its equilibrium point and resemble the community before the fire. This equilibrium state is referred to as the **climax community**, which will remain until the next disturbance. The climax community is typically characteristic of a given climate and geology. Although the community in equilibrium looks the same once it is attained, the equilibrium is a dynamic one with constant changes in abundance and sometimes species identities. The return of a natural ecosystem after agricultural activities is also a well-documented secondary succession process.



Secondary succession is seen in an oak and hickory forest after a

forest fire. A sequence of the community present at three successive times at the same location is depicted.

Section Summary

Communities include all the different species living in a given area. The variety of these species is referred to as biodiversity. Many organisms have developed defenses against predation and herbivory, including mechanical defenses, warning coloration, and mimicry. Two species cannot exist indefinitely in the same habitat competing directly for the same resources. Species may form symbiotic relationships such as commensalism, mutualism, or parasitism. Community structure is described by its foundation and keystone species. Communities respond to environmental disturbances by succession: the predictable appearance of different types of plant species, until a stable community structure is established.

Multiple Choice

Exercise:

Problem:

The first species to live on new land, such as that formed from volcanic lava, are called_________.

a. climax community
b. keystone species
c. foundation species
d. pioneer species

Solution:

D

Exercise:

Problem:

A symbiotic relationship where both of the co-existing species benefit from the interaction is called ________.

a. commensalism
b. parasitism
c. mutualism
d. communism

Solution:

C

Exercise:

Problem:

When an invasive species alters the community structure it is introduced to, what can the consequence be?

a. extinction of economically important species
b. reduced predation on some native species
c. increased predation on some native species
d. all of the above

Solution:

D

Free Response

Exercise:

Problem:

Describe the competitive exclusion principle and its effects on competing species.

Solution:

The competitive exclusion principles states that no two species competing for the same resources at the same time and place can co-exist over time. Thus, one of the competing species will eventually dominate. On the other hand, if the species evolve such that they use resources from different parts of the habitat or at different times of day, the two species can exist together indefinitely.

Exercise:

Problem:

Describe the potential effects when a keystone species is removed from a community.

Solution:

Removing a keystone species will have dramatic effects on the abundance of individuals in other populations, increasing some and decreasing others. This affects the interactions between populations such as competition and predator-prey relationships. In addition, the community may show a loss of diversity.

Glossary

climax community
: the final stage of succession, where a stable community is formed by a characteristic assortment of plant and animal species

competitive exclusion principle
: no two species within a habitat can coexist indefinitely when they compete for the same resources at the same time and place

environmental disturbance
: a change in the environment caused by natural disasters or human activities

foundation species
: a species which often forms the major structural portion of the habitat

host
: an organism a parasite lives on

island biogeography
: the study of life on island chains and how their geography interacts with the diversity of species found there

keystone species
: a species whose presence is key to maintaining biodiversity in an ecosystem and to upholding an ecological community’s structure

mimicry
: an adaptation in which an organism looks like another organism that is dangerous, toxic, or distasteful to its predators

mutualism
: a symbiotic relationship between two species where both species benefit

parasite
: an organism that uses resources from another species: the host

pioneer species
: the first species to appear in primary and secondary succession

primary succession
: the succession on land that previously has had no life

relative species abundance
: the absolute population size of a particular species relative to the population size of other species within the community

secondary succession
: the succession in response to environmental disturbances that move a community away from its equilibrium

species richness
: the number of different species in a community

Introduction
class="introduction"

The (a) Karner blue butterfly and (b) wild lupine live in oak-pine barren habitats in North America. (credit a: modification of work by John & Karen Hollingsworth, USFWS)

(a) (b)

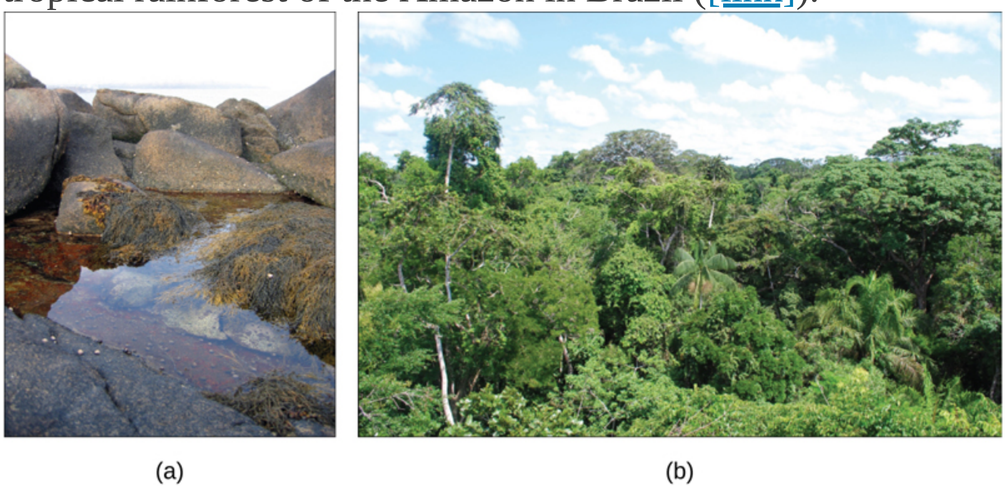
Ecosystem ecology is an extension of organismal, population, and community ecology. The ecosystem comprises all the biotic components (living things) and abiotic components (non-living things) in a particular geographic area. Some of the abiotic components include air, water, soil, and climate. Ecosystem biologists study how nutrients and energy are stored and moved among organisms and the surrounding atmosphere, soil, and water.

Wild lupine and Karner blue butterflies live in an oak-pine barren habitat in portions of Indiana, Michigan, Minnesota, Wisconsin, and New York ([link]). This habitat is characterized by natural disturbance in the form of fire and nutrient-poor soils that are low in nitrogen—important factors in the distribution of the plants that live in this habitat. Researchers interested in ecosystem ecology study the importance of limited resources in this ecosystem and the movement of resources (such as nutrients) through the biotic and abiotic portions of the ecosystem. Researchers also examine how organisms have adapted to their ecosystem.

Energy Flow through Ecosystems
By the end of this section, you will be able to:

- Describe the basic types of ecosystems on Earth
- Differentiate between food chains and food webs and recognize the importance of each
- Describe how organisms acquire energy in a food web and in associated food chains
- Explain how the efficiency of energy transfers between trophic levels effects ecosystem

An **ecosystem** is a community of living organisms and their abiotic (non-living) environment. Ecosystems can be small, such as the tide pools found near the rocky shores of many oceans, or large, such as those found in the tropical rainforest of the Amazon in Brazil ([link]).



A (a) tidal pool ecosystem in Matinicus Island, Maine, is a small ecosystem, while the (b) Amazon rainforest in Brazil is a large ecosystem. (credit a: modification of work by Jim Kuhn; credit b: modification of work by Ivan Mlinaric)

There are three broad categories of ecosystems based on their general environment: freshwater, marine, and terrestrial. Within these three

categories are individual ecosystem types based on the environmental habitat and organisms present.

Ecology of Ecosystems

Life in an ecosystem often involves competition for limited resources, which occurs both within a single species and between different species. Organisms compete for food, water, sunlight, space, and mineral nutrients. These resources provide the energy for metabolic processes and the matter to make up organisms' physical structures. Other critical factors influencing community dynamics are the components of its physical environment: a habitat's climate (seasons, sunlight, and rainfall), elevation, and geology. These can all be important environmental variables that determine which organisms can exist within a particular area.

Freshwater ecosystems are the least common, occurring on only 1.8 percent of Earth's surface. These systems comprise lakes, rivers, streams, and springs; they are quite diverse, and support a variety of animals, plants, fungi, protists and prokaryotes.

Marine ecosystems are the most common, comprising 75 percent of Earth's surface and consisting of three basic types: shallow ocean, deep ocean water, and deep ocean bottom. Shallow ocean ecosystems include extremely biodiverse coral reef ecosystems, yet the deep ocean water is known for large numbers of plankton and krill (small crustaceans) that support it. These two environments are especially important to aerobic respirators worldwide, as the phytoplankton perform 40 percent of all photosynthesis on Earth. Although not as diverse as the other two, deep ocean bottom ecosystems contain a wide variety of marine organisms. Such ecosystems exist even at depths where light is unable to penetrate through the water.

Terrestrial ecosystems, also known for their diversity, are grouped into large categories called biomes. A **biome** is a large-scale community of organisms, primarily defined on land by the dominant plant types that exist in geographic regions of the planet with similar climatic conditions. Examples of biomes include tropical rainforests, savannas, deserts, grasslands, temperate forests, and tundras. Grouping these ecosystems into

just a few biome categories obscures the great diversity of the individual ecosystems within them. For example, the saguaro cacti (*Carnegiea gigantean*) and other plant life in the Sonoran Desert, in the United States, are relatively diverse compared with the desolate rocky desert of Boa Vista, an island off the coast of Western Africa ([link]).

Desert ecosystems, like all ecosystems, can vary greatly. The desert in (a) Saguaro National Park, Arizona, has abundant plant life, while the rocky desert of (b) Boa Vista island, Cape Verde, Africa, is devoid of plant life. (credit a: modification of work by Jay Galvin; credit b: modification of work by Ingo Wölbern)

Ecosystems and Disturbance

Ecosystems are complex with many interacting parts. They are routinely exposed to various disturbances: changes in the environment that affect their compositions, such as yearly variations in rainfall and temperature. Many disturbances are a result of natural processes. For example, when lightning causes a forest fire and destroys part of a forest ecosystem, the ground is eventually populated with grasses, followed by bushes and shrubs, and later mature trees: thus, the forest is restored to its former state.

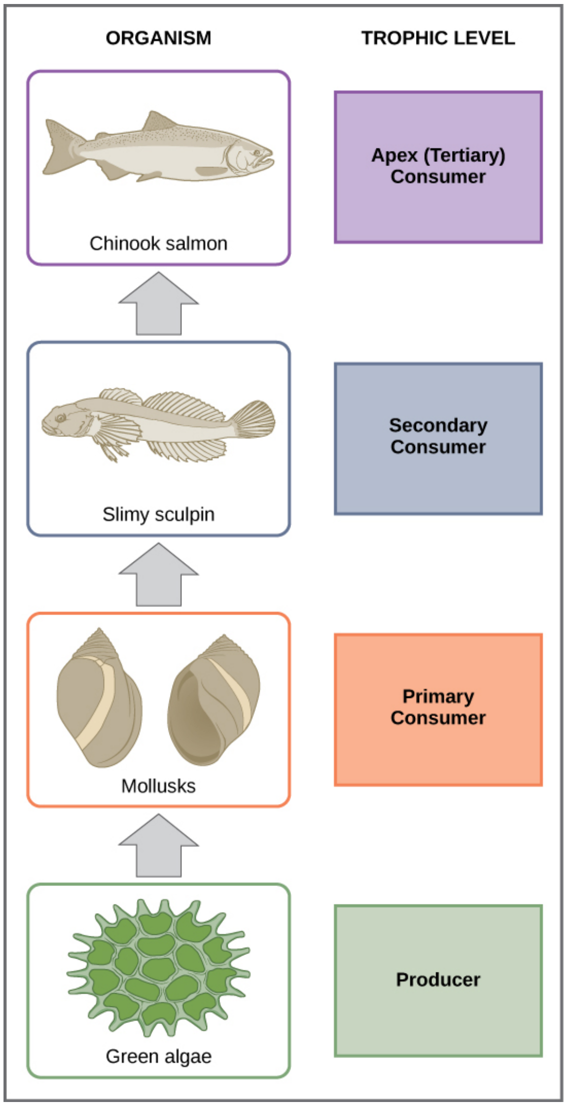
This process is so universal that ecologists have given it a name—succession. The impact of environmental disturbances caused by human activities is now as significant as the changes wrought by natural processes. Human agricultural practices, air pollution, acid rain, global deforestation, overfishing, oil spills, and illegal dumping on land and into the ocean all have impacts on ecosystems.

Equilibrium is a dynamic state of an ecosystem in which, despite changes in species numbers and occurrence, biodiversity remains somewhat constant. In ecology, two parameters are used to measure changes in ecosystems: resistance and resilience. The ability of an ecosystem to remain at equilibrium in spite of disturbances is called **resistance**. The speed at which an ecosystem recovers equilibrium after being disturbed is called **resilience**. Ecosystem resistance and resilience are especially important when considering human impact. The nature of an ecosystem may change to such a degree that it can lose its resilience entirely. This process can lead to the complete destruction or irreversible altering of the ecosystem.

Food Chains and Food Webs

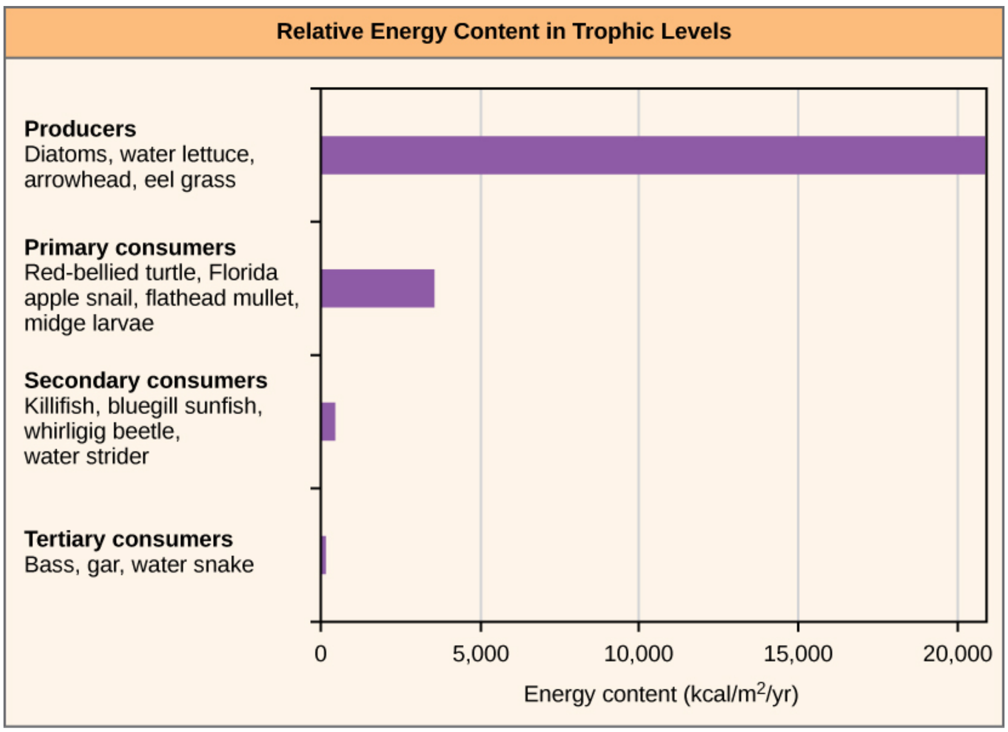
A **food chain** is a linear sequence of organisms through which nutrients and energy pass as one organism eats another; the levels in the food chain are producers, primary consumers, higher-level consumers, and finally decomposers. These levels are used to describe ecosystem structure and dynamics. There is a single path through a food chain. Each organism in a food chain occupies a specific **trophic level** (energy level), its position in the food chain or food web.

In many ecosystems, the base, or foundation, of the food chain consists of photosynthetic organisms (plants or phytoplankton), which are called **producers**. The organisms that consume the producers are herbivores: the **primary consumers**. **Secondary consumers** are usually carnivores that eat the primary consumers. **Tertiary consumers** are carnivores that eat other carnivores. Higher-level consumers feed on the next lower trophic levels, and so on, up to the organisms at the top of the food chain: the **apex consumers**. In the Lake Ontario food chain, shown in [link], the Chinook salmon is the apex consumer at the top of this food chain.



These are the trophic levels of a food chain in Lake Ontario at the United States–Canada border. Energy and nutrients flow from photosynthetic green algae at the base to the top of the food chain: the Chinook salmon. (credit: modification of work by National Oceanic and Atmospheric Administration/NOAA)

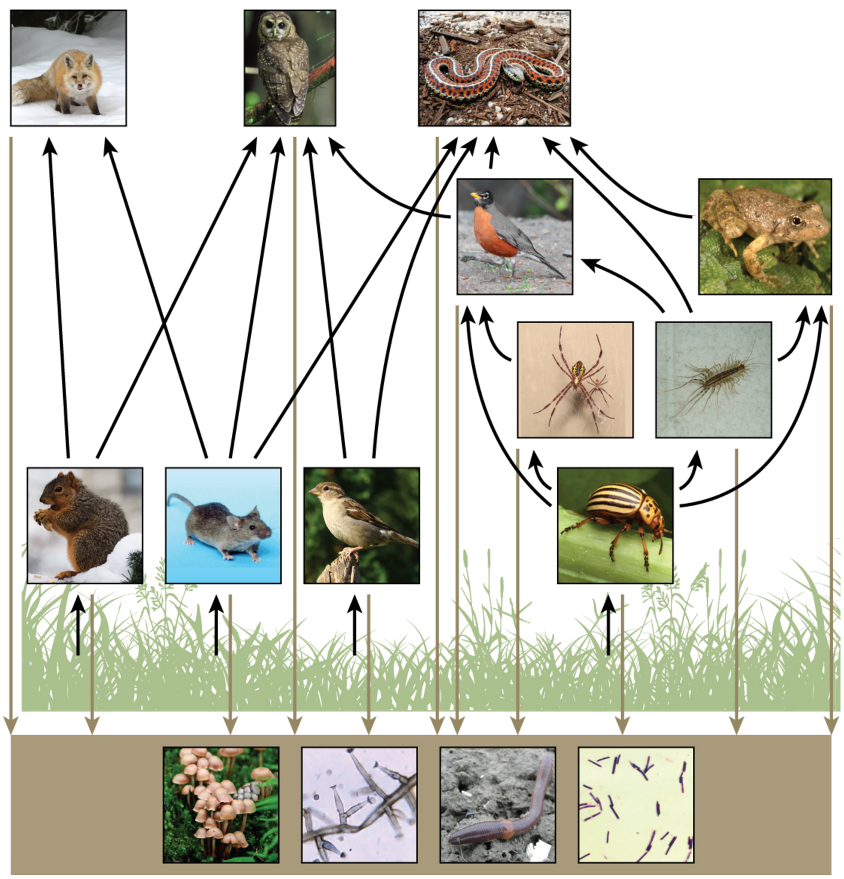
One major factor that limits the number of steps in a food chain is energy. Energy is lost at each trophic level and between trophic levels as heat and in the transfer to decomposers ([link]). Thus, after a limited number of trophic energy transfers, the amount of energy remaining in the food chain may not be great enough to support viable populations at yet a higher trophic level.



The relative energy in trophic levels in a Silver Springs, Florida, ecosystem is shown. Each trophic level has less energy available, and usually, but not always, supports a smaller mass of organisms at the next level.

There is a one problem when using food chains to describe most ecosystems. Even when all organisms are grouped into appropriate trophic levels, some of these organisms can feed on more than one trophic level;

likewise, some of these organisms can also be fed on from multiple trophic levels. In addition, species feed on and are eaten by more than one species. In other words, the linear model of ecosystems, the food chain, is a hypothetical, overly simplistic representation of ecosystem structure. A holistic model—which includes all the interactions between different species and their complex interconnected relationships with each other and with the environment—is a more accurate and descriptive model for ecosystems. A **food web** is a concept that accounts for the multiple trophic (feeding) interactions between each species and the many species it may feed on, or that feed on it. In a food web, the several trophic connections between each species and the other species that interact with it may cross multiple trophic levels. The matter and energy movements of virtually all ecosystems are more accurately described by food webs ([link]).



This food web shows the interactions between

organisms across trophic levels. Arrows point from an organism that is consumed to the organism that consumes it. All the producers and consumers eventually become nourishment for the decomposers (fungi, mold, earthworms, and bacteria in the soil). (credit "fox": modification of work by Kevin Bacher, NPS; credit "owl": modification of work by John and Karen Hollingsworth, USFWS; credit "snake": modification of work by Steve Jurvetson; credit "robin": modification of work by Alan Vernon; credit "frog": modification of work by Alessandro Catenazzi; credit "spider": modification of work by "Sanba38"/Wikimedia Commons; credit "centipede": modification of work by “Bauerph”/Wikimedia Commons; credit "squirrel": modification of work by Dawn Huczek; credit "mouse": modification of work by NIGMS, NIH; credit "sparrow": modification of work by David Friel; credit "beetle": modification of work by Scott Bauer, USDA Agricultural Research Service; credit "mushrooms": modification of work by Chris Wee; credit "mold": modification of work by Dr. Lucille Georg, CDC; credit "earthworm": modification of work by Rob Hille; credit "bacteria": modification of work by Don Stalons, CDC)

Note:
Concept in Action

Head to this online interactive simulator to investigate food web function. In the *Interactive Labs* box, under Food Web, click **Step 1**. Read the instructions first, and then click **Step 2** for additional instructions. When you are ready to create a simulation, in the upper-right corner of the *Interactive Labs* box, click **OPEN SIMULATOR**.

Two general types of food webs are often shown interacting within a single ecosystem. A **grazing food web** has plants or other photosynthetic organisms at its base, followed by herbivores and various carnivores. A **detrital food web** consists of a base of organisms that feed on decaying organic matter (dead organisms), including decomposers (which break down dead and decaying organisms) and detritivores (which consume organic detritus). These organisms are usually bacteria, fungi, and invertebrate animals that recycle organic material back into the biotic part of the ecosystem as they themselves are consumed by other organisms. As ecosystems require a method to recycle material from dead organisms, grazing food webs have an associated detrital food web. For example, in a meadow ecosystem, plants may support a grazing food web of different organisms, primary and other levels of consumers, while at the same time supporting a detrital food web of bacteria and fungi feeding off dead plants and animals. Simultaneously, a detrital food web can contribute energy to a grazing food web, as when a robin eats an earthworm.

How Organisms Acquire Energy in a Food Web

All living things require energy in one form or another. Energy is used by most complex metabolic pathways (usually in the form of ATP), especially those responsible for building large molecules from smaller compounds. Living organisms would not be able to assemble macromolecules (proteins,

lipids, nucleic acids, and complex carbohydrates) from their monomers without a constant energy input.

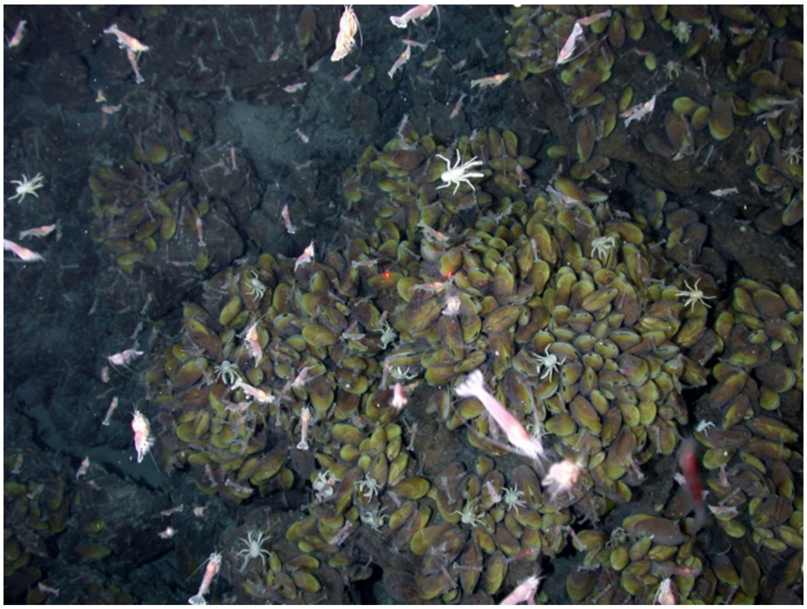
Food-web diagrams illustrate how energy flows directionally through ecosystems. They can also indicate how efficiently organisms acquire energy, use it, and how much remains for use by other organisms of the food web. Energy is acquired by living things in two ways: autotrophs harness light or chemical energy and heterotrophs acquire energy through the consumption and digestion of other living or previously living organisms.

Photosynthetic and chemosynthetic organisms are **autotrophs**, which are organisms capable of synthesizing their own food (more specifically, capable of using inorganic carbon as a carbon source). Photosynthetic autotrophs (**photoautotrophs**) use sunlight as an energy source, and chemosynthetic autotrophs (**chemoautotrophs**) use inorganic molecules as an energy source. Autotrophs are critical for most ecosystems: they are the producer trophic level. Without these organisms, energy would not be available to other living organisms, and life itself would not be possible.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, are the energy source for a majority of the world’s ecosystems. These ecosystems are often described by grazing and detrital food webs. Photoautotrophs harness the Sun’s solar energy by converting it to chemical energy in the form of ATP (and NADP). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose. The rate at which photosynthetic producers incorporate energy from the Sun is called **gross primary productivity**. However, not all of the energy incorporated by producers is available to the other organisms in the food web because producers must also grow and reproduce, which consumes energy. **Net primary productivity** is the energy that remains in the producers after accounting for these organisms’ respiration and heat loss. The net productivity is then available to the primary consumers at the next trophic level.

Chemoautotrophs are primarily bacteria and archaea that are found in rare ecosystems where sunlight is not available, such as those associated with dark caves or hydrothermal vents at the bottom of the ocean ([link]). Many

chemoautotrophs in hydrothermal vents use hydrogen sulfide (H_2S), which is released from the vents as a source of chemical energy; this allows them to synthesize complex organic molecules, such as glucose, for their own energy and, in turn, supplies energy to the rest of the ecosystem.

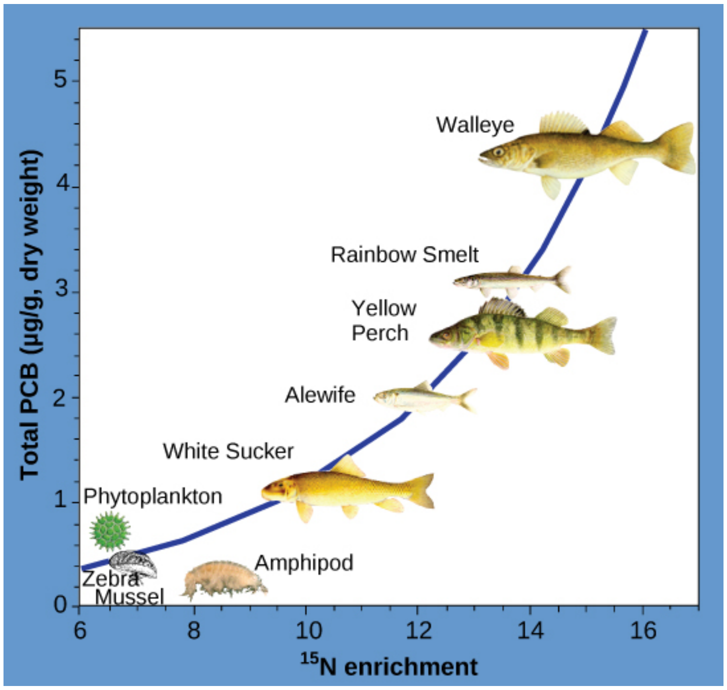


Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean’s surface. This picture was taken in 2006 at the submerged NW Eifuku volcano off the coast of Japan by the National Oceanic and Atmospheric Administration (NOAA). The summit of this highly active volcano lies 1535 m below the surface.

Consequences of Food Webs: Biological Magnification

One of the most important consequences of ecosystem dynamics in terms of human impact is biomagnification. **Biomagnification** is the increasing concentration of persistent, toxic substances in organisms at each successive trophic level. These are substances that are fat soluble, not water soluble, and are stored in the fat reserves of each organism. Many substances have been shown to biomagnify, including classical studies with the pesticide dichlorodiphenyltrichloroethane (DDT), which were described in the 1960s bestseller, *Silent Spring* by Rachel Carson. DDT was a commonly used pesticide before its dangers to apex consumers, such as the bald eagle, became known. In aquatic ecosystems, organisms from each trophic level consumed many organisms in the lower level, which caused DDT to increase in birds (apex consumers) that ate fish. Thus, the birds accumulated sufficient amounts of DDT to cause fragility in their eggshells. This effect increased egg breakage during nesting and was shown to have devastating effects on these bird populations. The use of DDT was banned in the United States in the 1970s.

Other substances that biomagnify are polychlorinated biphenyls (PCB), which were used as coolant liquids in the United States until their use was banned in 1979, and heavy metals, such as mercury, lead, and cadmium. These substances are best studied in aquatic ecosystems, where predatory fish species accumulate very high concentrations of toxic substances that are at quite low concentrations in the environment and in producers. As illustrated in a study performed by the NOAA in the Saginaw Bay of Lake Huron of the North American Great Lakes ([link]), PCB concentrations increased from the producers of the ecosystem (phytoplankton) through the different trophic levels of fish species. The apex consumer, the walleye, has more than four times the amount of PCBs compared to phytoplankton. Also, based on results from other studies, birds that eat these fish may have PCB levels at least one order of magnitude higher than those found in the lake fish.



This chart shows the PCB concentrations found at the various trophic levels in the Saginaw Bay ecosystem of Lake Huron. Notice that the fish in the higher trophic levels accumulate more PCBs than those in lower trophic levels. (credit: Patricia Van Hoof, NOAA)

Other concerns have been raised by the biomagnification of heavy metals, such as mercury and cadmium, in certain types of seafood. The United States Environmental Protection Agency recommends that pregnant women and young children should not consume any swordfish, shark, king mackerel, or tilefish because of their high mercury content. These individuals are advised to eat fish low in mercury: salmon, shrimp, pollock, and catfish. Biomagnification is a good example of how ecosystem dynamics can affect our everyday lives, even influencing the food we eat.

Section Summary

Ecosystems exist underground, on land, at sea, and in the air. Organisms in an ecosystem acquire energy in a variety of ways, which is transferred between trophic levels as the energy flows from the base to the top of the food web, with energy being lost at each transfer. There is energy lost at each trophic level, so the lengths of food chains are limited because there is a point where not enough energy remains to support a population of consumers. Fat soluble compounds biomagnify up a food chain causing damage to top consumers. even when environmental concentrations of a toxin are low.

Multiple Choice

Exercise:

Problem: Decomposers are associated with which class of food web?

a. grazing
b. detrital
c. inverted
d. aquatic

Solution:

B

Exercise:

Problem:

The producer in an ocean grazing food web is usually a _________.

a. plant
b. animal
c. fungi
d. plankton

Solution:

D

Exercise:

Problem:

Which term describes the process whereby toxic substances increase along trophic levels of an ecosystem?

a. biomassification
b. biomagnification
c. bioentropy
d. heterotrophy

Solution:

B

Free Response

Exercise:

Problem:

Compare grazing and detrital food webs. Why would they both be present in the same ecosystem?

Solution:

Grazing food webs have a producer at their base, which is either a plant for terrestrial ecosystems or a phytoplankton for aquatic ecosystems. The producers pass their energy to the various trophic levels of consumers. At the base of detrital food webs are the decomposers, which pass their energy to a variety of other consumers. Detrital food webs are important for the health of many grazing food

webs because they eliminate dead and decaying organic material, thus clearing space for new organisms and removing potential causes of disease.

Glossary

autotroph
: an organism capable of synthesizing its own food molecules from smaller inorganic molecules

apex consumer
: an organism at the top of the food chain

biomagnification
: an increasing concentration of persistent, toxic substances in organisms at each trophic level, from the producers to the apex consumers

biome
: a large-scale community of organisms, primarily defined on land by the dominant plant types that exist in geographic regions of the planet with similar climatic conditions

chemoautotroph
: an organism capable of synthesizing its own food using energy from inorganic molecules

detrital food web
: a type of food web that is supported by dead or decaying organisms rather than by living autotrophs; these are often associated with grazing food webs within the same ecosystem

ecosystem
: a community of living organisms and their interactions with their abiotic environment

equilibrium

the steady state of a system in which the relationships between elements of the system do not change

food chain
: a linear sequence of trophic (feeding) relationships of producers, primary consumers, and higher level consumers

food web
: a web of trophic (feeding) relationships among producers, primary consumers, and higher level consumers in an ecosystem

grazing food web
: a type of food web in which the producers are either plants on land or phytoplankton in the water; often associated with a detrital food web within the same ecosystem

gross primary productivity
: the rate at which photosynthetic producers incorporate energy from the Sun

net primary productivity
: the energy that remains in the producers after accounting for the organisms' respiration and heat loss

photoautotroph
: an organism that uses sunlight as an energy source to synthesize its own food molecules

primary consumer
: the trophic level that obtains its energy from the producers of an ecosystem

producer
: the trophic level that obtains its energy from sunlight, inorganic chemicals, or dead or decaying organic material

resilience (ecological)

the speed at which an ecosystem recovers equilibrium after being disturbed

resistance (ecological)
: the ability of an ecosystem to remain at equilibrium in spite of disturbances

secondary consumer
: a trophic level in an ecosystem, usually a carnivore that eats a primary consumer

tertiary consumer
: a trophic level in an ecosystem, usually carnivores that eat other carnivores

trophic level
: the position of a species or group of species in a food chain or a food web

Biogeochemical Cycles
By the end of this section, you will be able to:

- Discuss the biogeochemical cycles of water, carbon, nitrogen, phosphorus, and sulfur
- Explain how human activities have impacted these cycles and the resulting potential consequences for Earth

Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the transfers between trophic levels. Rather than flowing through an ecosystem, the matter that makes up living organisms is conserved and recycled. The six most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, phosphorus, and sulfur—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath Earth's surface. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in the cycling of elements on Earth. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their nonliving environment is called a **biogeochemical cycle**.

Water, which contains hydrogen and oxygen, is essential to all living processes. The **hydrosphere** is the area of Earth where water movement and storage occurs: as liquid water on the surface (rivers, lakes, oceans) and beneath the surface (groundwater) or ice, (polar ice caps and glaciers), and as water vapor in the atmosphere. Carbon is found in all organic macromolecules and is an important constituent of fossil fuels. Nitrogen is a major component of our nucleic acids and proteins and is critical to human agriculture. Phosphorus, a major component of nucleic acids, is one of the main ingredients (along with nitrogen) in artificial fertilizers used in agriculture, which has environmental impacts on our surface water. Sulfur, critical to the three-dimensional folding of proteins (as in disulfide binding), is released into the atmosphere by the burning of fossil fuels.

The cycling of these elements is interconnected. For example, the movement of water is critical for the leaching of nitrogen and phosphate into rivers, lakes, and oceans. The ocean is also a major reservoir for

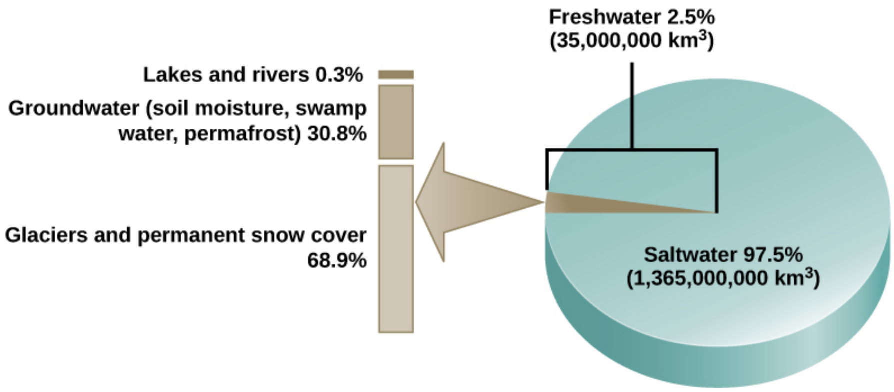
carbon. Thus, mineral nutrients are cycled, either rapidly or slowly, through the entire biosphere between the biotic and abiotic world and from one living organism to another.

Note:
Concept in Action

Head to this website to learn more about biogeochemical cycles.

The Water Cycle

Water is essential for all living processes. The human body is more than one-half water and human cells are more than 70 percent water. Thus, most land animals need a supply of fresh water to survive. Of the stores of water on Earth, 97.5 percent is salt water ([link]). Of the remaining water, 99 percent is locked as underground water or ice. Thus, less than one percent of fresh water is present in lakes and rivers. Many living things are dependent on this small amount of surface fresh water supply, a lack of which can have important effects on ecosystem dynamics. Humans, of course, have developed technologies to increase water availability, such as digging wells to harvest groundwater, storing rainwater, and using desalination to obtain drinkable water from the ocean. Although this pursuit of drinkable water has been ongoing throughout human history, the supply of fresh water continues to be a major issue in modern times.



Only 2.5 percent of water on Earth is fresh water, and less than 1 percent of fresh water is easily accessible to living things.

The various processes that occur during the cycling of water are illustrated in [link]. The processes include the following:

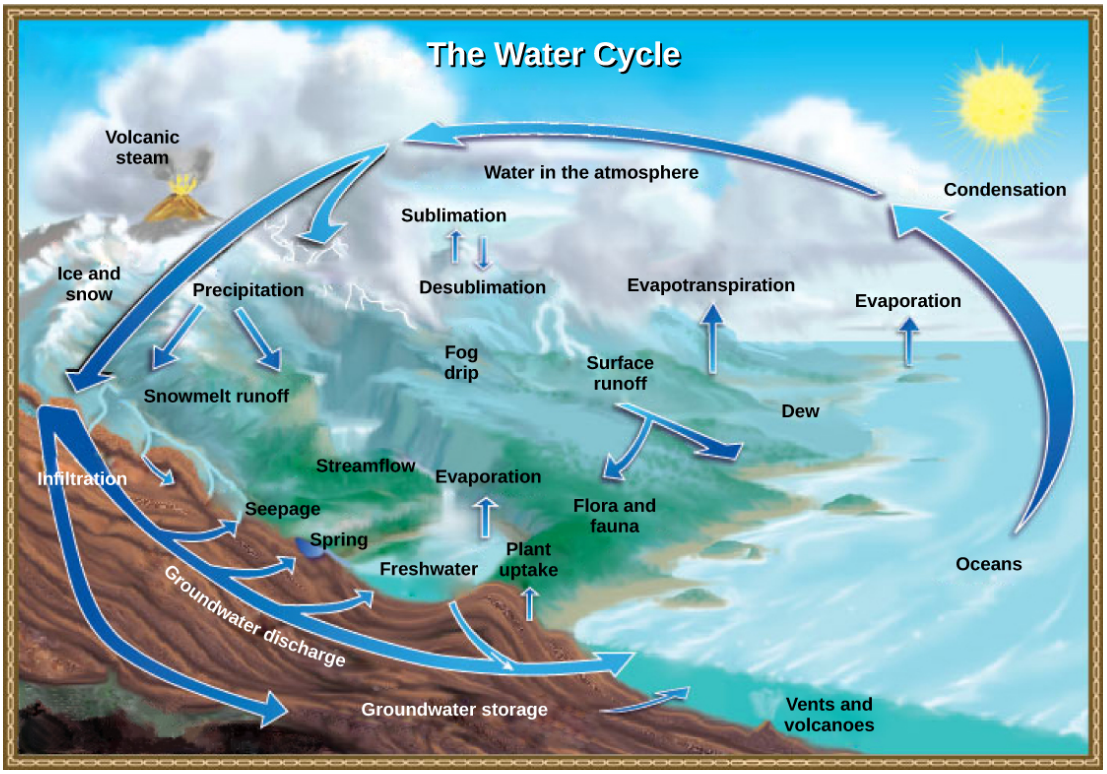
- evaporation and sublimation
- condensation and precipitation
- subsurface water flow
- surface runoff and snowmelt
- streamflow

The water cycle is driven by the Sun’s energy as it warms the oceans and other surface waters. This leads to evaporation (water to water vapor) of liquid surface water and sublimation (ice to water vapor) of frozen water, thus moving large amounts of water into the atmosphere as water vapor. Over time, this water vapor condenses into clouds as liquid or frozen droplets and eventually leads to precipitation (rain or snow), which returns water to Earth’s surface. Rain reaching Earth’s surface may evaporate again, flow over the surface, or percolate into the ground. Most easily observed is surface runoff: the flow of fresh water either from rain or melting ice. Runoff can make its way through streams and lakes to the oceans or flow directly to the oceans themselves.

In most natural terrestrial environments rain encounters vegetation before it reaches the soil surface. A significant percentage of water evaporates immediately from the surfaces of plants. What is left reaches the soil and begins to move down. Surface runoff will occur only if the soil becomes saturated with water in a heavy rainfall. Most water in the soil will be taken up by plant roots. The plant will use some of this water for its own metabolism, and some of that will find its way into animals that eat the plants, but much of it will be lost back to the atmosphere through a process known as evapotranspiration. Water enters the vascular system of the plant through the roots and evaporates, or transpires, through the stomata of the leaves. Water in the soil that is not taken up by a plant and that does not evaporate is able to percolate into the subsoil and bedrock. Here it forms groundwater.

Groundwater is a significant reservoir of fresh water. It exists in the pores between particles in sand and gravel, or in the fissures in rocks. Shallow groundwater flows slowly through these pores and fissures and eventually finds its way to a stream or lake where it becomes a part of the surface water again. Streams do not flow because they are replenished from rainwater directly; they flow because there is a constant inflow from groundwater below. Some groundwater is found very deep in the bedrock and can persist there for millennia. Most groundwater reservoirs, or aquifers, are the source of drinking or irrigation water drawn up through wells. In many cases these aquifers are being depleted faster than they are being replenished by water percolating down from above.

Rain and surface runoff are major ways in which minerals, including carbon, nitrogen, phosphorus, and sulfur, are cycled from land to water. The environmental effects of runoff will be discussed later as these cycles are described.



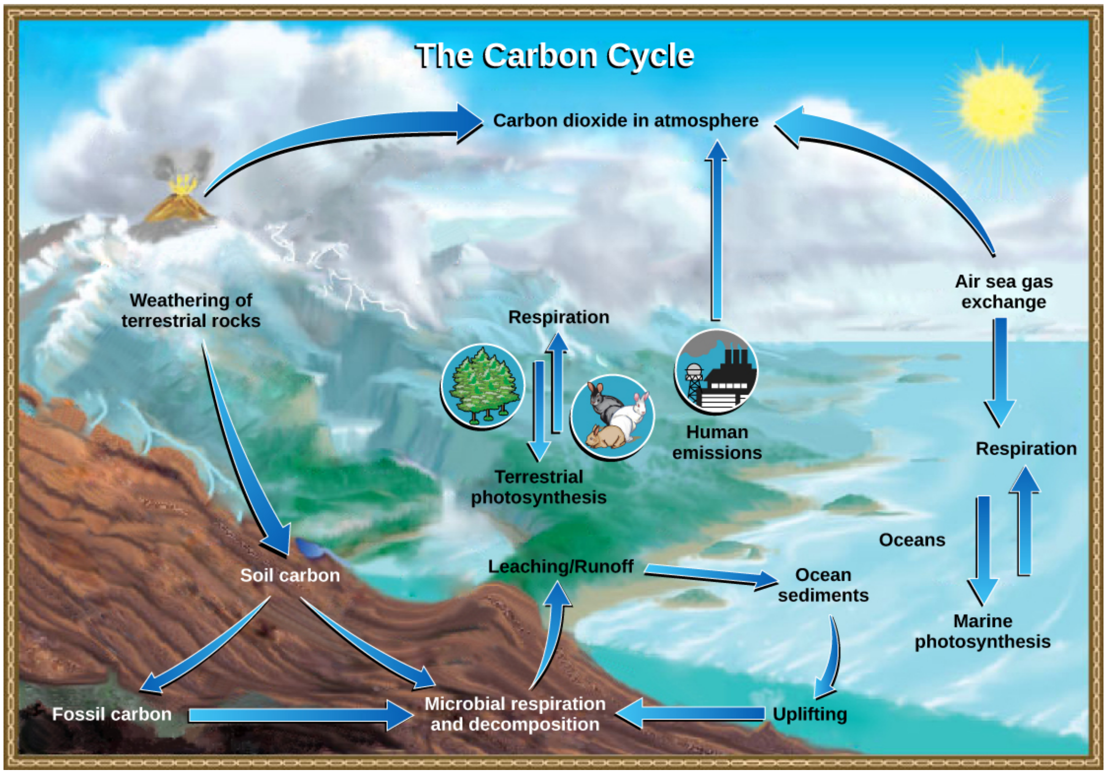
Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Carbon Cycle

Carbon is the fourth most abundant element in living organisms. Carbon is present in all organic molecules, and its role in the structure of macromolecules is of primary importance to living organisms. Carbon compounds contain energy, and many of these compounds from plants and algae have remained stored as fossilized carbon, which humans use as fuel. Since the 1800s, the use of fossil fuels has accelerated. As global demand for Earth's limited fossil fuel supplies has risen since the beginning of the Industrial Revolution, the amount of carbon dioxide in our atmosphere has

increased as the fuels are burned. This increase in carbon dioxide has been associated with climate change and is a major environmental concern worldwide.

The carbon cycle is most easily studied as two interconnected subcycles: one dealing with rapid carbon exchange among living organisms and the other dealing with the long-term cycling of carbon through geologic processes. The entire carbon cycle is shown in [link].



Carbon dioxide gas exists in the atmosphere and is dissolved in water. Photosynthesis converts carbon dioxide gas to organic carbon, and respiration cycles the organic carbon back into carbon dioxide gas. Long-term storage of organic carbon occurs when matter from living organisms is buried deep underground and becomes fossilized. Volcanic activity and, more recently, human emissions bring this stored carbon back into the carbon cycle. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Biological Carbon Cycle

Living organisms are connected in many ways, even between ecosystems. A good example of this connection is the exchange of carbon between heterotrophs and autotrophs within and between ecosystems by way of atmospheric carbon dioxide. Carbon dioxide is the basic building block that autotrophs use to build multi-carbon, high-energy compounds, such as glucose. The energy harnessed from the Sun is used by these organisms to form the covalent bonds that link carbon atoms together. These chemical bonds store this energy for later use in the process of respiration. Most terrestrial autotrophs obtain their carbon dioxide directly from the atmosphere, while marine autotrophs acquire it in the dissolved form (carbonic acid, HCO_3^-). However the carbon dioxide is acquired, a byproduct of fixing carbon in organic compounds is oxygen. Photosynthetic organisms are responsible for maintaining approximately 21 percent of the oxygen content of the atmosphere that we observe today.

The partners in biological carbon exchange are the heterotrophs (especially the primary consumers, largely herbivores). Heterotrophs acquire the high-energy carbon compounds from the autotrophs by consuming them and breaking them down by respiration to obtain cellular energy, such as ATP. The most efficient type of respiration, aerobic respiration, requires oxygen obtained from the atmosphere or dissolved in water. Thus, there is a constant exchange of oxygen and carbon dioxide between the autotrophs (which need the carbon) and the heterotrophs (which need the oxygen). Autotrophs also respire and consume the organic molecules they form: using oxygen and releasing carbon dioxide. They release more oxygen gas as a waste product of photosynthesis than they use for their own respiration; therefore, there is excess available for the respiration of other aerobic organisms. Gas exchange through the atmosphere and water is one way that the carbon cycle connects all living organisms on Earth.

The Biogeochemical Carbon Cycle

The movement of carbon through land, water, and air is complex, and, in many cases, it occurs much more slowly geologically than the movement between living organisms. Carbon is stored for long periods in what are known as carbon reservoirs, which include the atmosphere, bodies of liquid water (mostly oceans), ocean sediment, soil, rocks (including fossil fuels), and Earth's interior.

As stated, the atmosphere is a major reservoir of carbon in the form of carbon dioxide that is essential to the process of photosynthesis. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each, and each one affects the other reciprocally. Carbon dioxide (CO_2) from the atmosphere dissolves in water and, unlike oxygen and nitrogen gas, reacts with water molecules to form ionic compounds. Some of these ions combine with calcium ions in the seawater to form calcium carbonate ($CaCO_3$), a major component of the shells of marine organisms. These organisms eventually form sediments on the ocean floor. Over geologic time, the calcium carbonate forms limestone, which comprises the largest carbon reservoir on Earth.

On land, carbon is stored in soil as organic carbon as a result of the decomposition of living organisms or from weathering of terrestrial rock and minerals. Deeper under the ground, at land and at sea, are fossil fuels, the anaerobically decomposed remains of plants that take millions of years to form. Fossil fuels are considered a non-renewable resource because their use far exceeds their rate of formation. A **non-renewable resource** is either regenerated very slowly or not at all. Another way for carbon to enter the atmosphere is from land (including land beneath the surface of the ocean) by the eruption of volcanoes and other geothermal systems. Carbon sediments from the ocean floor are taken deep within Earth by the process of **subduction**: the movement of one tectonic plate beneath another. Carbon is released as carbon dioxide when a volcano erupts or from volcanic hydrothermal vents.

Carbon dioxide is also added to the atmosphere by the animal husbandry practices of humans. The large number of land animals raised to feed Earth's growing human population results in increased carbon-dioxide levels in the atmosphere caused by their respiration. This is another example of how human activity indirectly affects biogeochemical cycles in a significant way. Although much of the debate about the future effects of increasing atmospheric carbon on climate change focuses on fossils fuels, scientists take natural processes, such as volcanoes, plant growth, soil carbon levels, and respiration, into account as they model and predict the future impact of this increase.

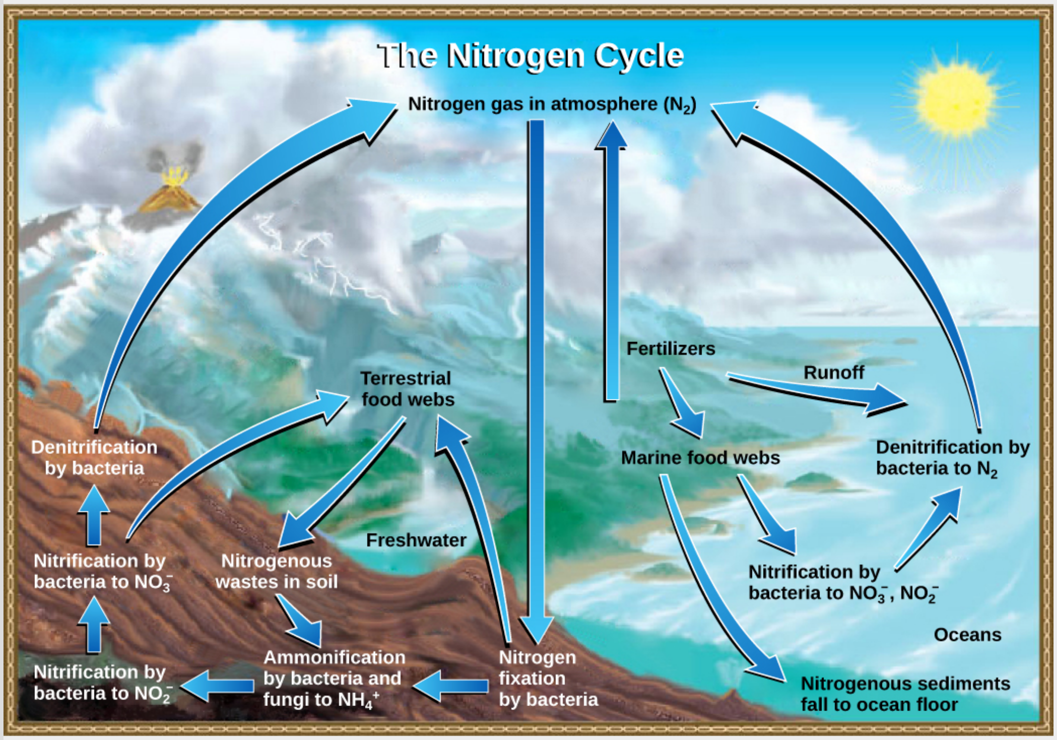
The Nitrogen Cycle

Getting nitrogen into the living world is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (which exists as tightly bonded, triple covalent N_2) even though this molecule comprises approximately 78 percent of the atmosphere. Nitrogen enters the living world via free-living and symbiotic bacteria, which incorporate nitrogen into their macromolecules through nitrogen fixation (conversion of N_2). Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Cyanobacteria are able to use inorganic sources of nitrogen to "fix" nitrogen. *Rhizobium* bacteria live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. Free-living bacteria, such as *Azotobacter*, are also important nitrogen fixers.

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. As shown in [link], the nitrogen that enters living systems by nitrogen fixation is eventually converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the ammonification process converts nitrogenous waste from living animals or from the remains of dead animals into ammonium (NH_4^+) by certain bacteria and fungi. Second, this ammonium is then converted to nitrites (NO_2^-) by nitrifying bacteria, such as *Nitrosomonas*, through nitrification. Subsequently, nitrites

are converted to nitrates (NO_3^-) by similar organisms. Lastly, the process of denitrification occurs, whereby bacteria, such as *Pseudomonas* and *Clostridium*, convert the nitrates into nitrogen gas, thus allowing it to re-enter the atmosphere.

Note:
Art Connection



Nitrogen enters the living world from the atmosphere through nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Which of the following statements about the nitrogen cycle is false?

a. Ammonification converts organic nitrogenous matter from living organisms into ammonium (NH_4^+).
b. Denitrification by bacteria converts nitrates (NO_3^-)to nitrogen gas (N_2).
c. Nitrification by bacteria converts nitrates (NO_3^-)to nitrites (NO_2^-)
d. Nitrogen fixing bacteria convert nitrogen gas (N_2) into organic compounds.

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers (which contain nitrogen and phosphorus compounds) in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Atmospheric nitrogen (other than N_2) is associated with several effects on Earth's ecosystems including the production of acid rain (as nitric acid, HNO_3) and greenhouse gas effects (as nitrous oxide, N_2O), potentially causing climate change. A major effect from fertilizer runoff is saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the overgrowth of algae and a number of consequential problems.

A similar process occurs in the marine nitrogen cycle, where the ammonification, nitrification, and denitrification processes are performed by marine bacteria and archaea. Some of this nitrogen falls to the ocean floor as sediment, which can then be moved to land in geologic time by uplift of Earth's surface, and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle. [footnote]

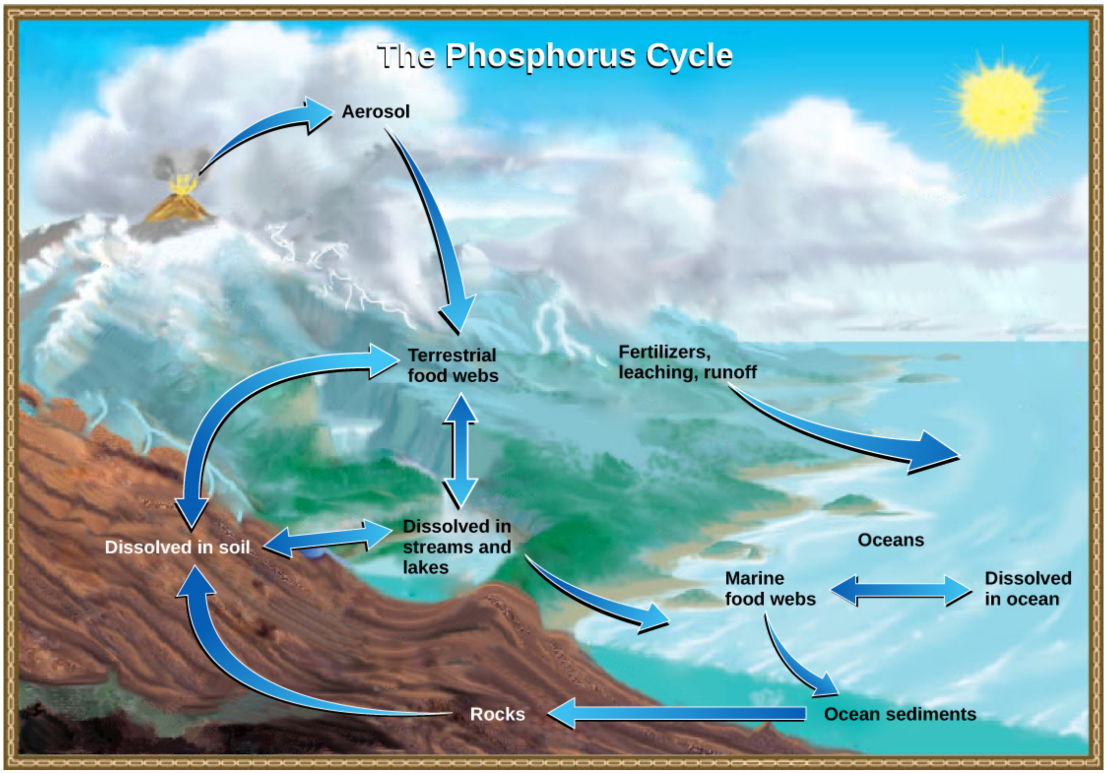
Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, "Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock," *Nature* 477, no. 7362 (2011): 78–81.

The Phosphorus Cycle

Phosphorus is an essential nutrient for living processes; it is a major component of nucleic acids and phospholipids, and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic, particularly freshwater, ecosystems.

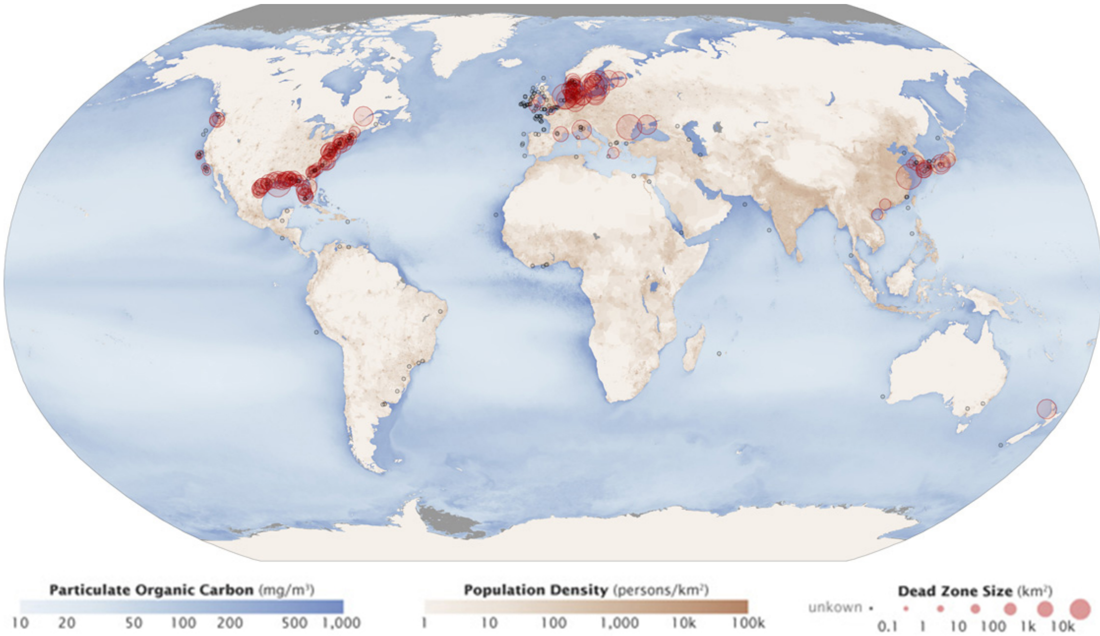
Phosphorus occurs in nature as the phosphate ion (PO_4^{3-}). In addition to phosphate runoff as a result of human activity, natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean. This rock has its origins in the ocean. Phosphate-containing ocean sediments form primarily from the bodies of ocean organisms and from their excretions. However, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is moved to land over geologic time by the uplifting of Earth's surface. ([link])

Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine organisms. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion having an oceanic residence time between 20,000 and 100,000 years.



In nature, phosphorus exists as the phosphate ion (PO_4^{3-}). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans in surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Excess phosphorus and nitrogen that enter these ecosystems from fertilizer runoff and from sewage cause excessive growth of algae. The subsequent death and decay of these organisms depletes dissolved oxygen, which leads to the death of aquatic organisms, such as shellfish and finfish. This process is responsible for dead zones in lakes and at the mouths of many major rivers and for massive fish kills, which often occur during the summer months (see [link]).



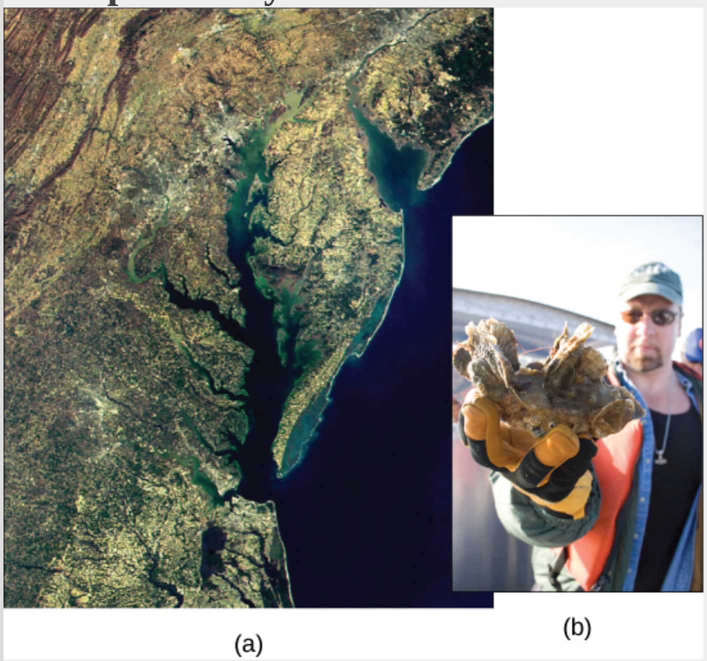
Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. Worldwide, large dead zones are found in areas of high population density. (credit: Robert Simmon, Jesse Allen, NASA Earth Observatory)

A **dead zone** is an area in lakes and oceans near the mouths of rivers where large areas are periodically depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping toxic chemicals, and other human activities. The number of dead zones has increased for several years, and more than 400 of these zones were present as of 2008. One of the worst dead zones is off the coast of the United States in the Gulf of Mexico: fertilizer runoff from the Mississippi River basin created a dead zone of over 8,463 square miles. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

Note:

Careers in Action

Chesapeake Bay



This (a) satellite image shows the Chesapeake Bay, an ecosystem affected by phosphate and nitrate runoff. A (b) member of the Army Corps of Engineers holds a clump of oysters being used as a part of the oyster restoration effort in the bay. (credit a: modification of work by NASA/MODIS; credit b: modification of work by U.S. Army)

The Chesapeake Bay ([link]**a**) is one of the most scenic areas on Earth; it is now in distress and is recognized as a case study of a declining ecosystem. In the 1970s, the Chesapeake Bay was one of the first aquatic ecosystems to have identified dead zones, which continue to kill many fish and bottom-dwelling species such as clams, oysters, and worms. Several species have declined in the Chesapeake Bay because surface water runoff contains excess nutrients from artificial fertilizer use on land. The source of the fertilizers (with high nitrogen and phosphate content) is not limited

to agricultural practices. There are many nearby urban areas and more than 150 rivers and streams empty into the bay that are carrying fertilizer runoff from lawns and gardens. Thus, the decline of the Chesapeake Bay is a complex issue and requires the cooperation of industry, agriculture, and individual homeowners.

Of particular interest to conservationists is the oyster population ([link]**b**); it is estimated that more than 200,000 acres of oyster reefs existed in the bay in the 1700s, but that number has now declined to only 36,000 acres. Oyster harvesting was once a major industry for Chesapeake Bay, but it declined 88 percent between 1982 and 2007. This decline was caused not only by fertilizer runoff and dead zones, but also because of overharvesting. Oysters require a certain minimum population density because they must be in close proximity to reproduce. Human activity has altered the oyster population and locations, thus greatly disrupting the ecosystem.

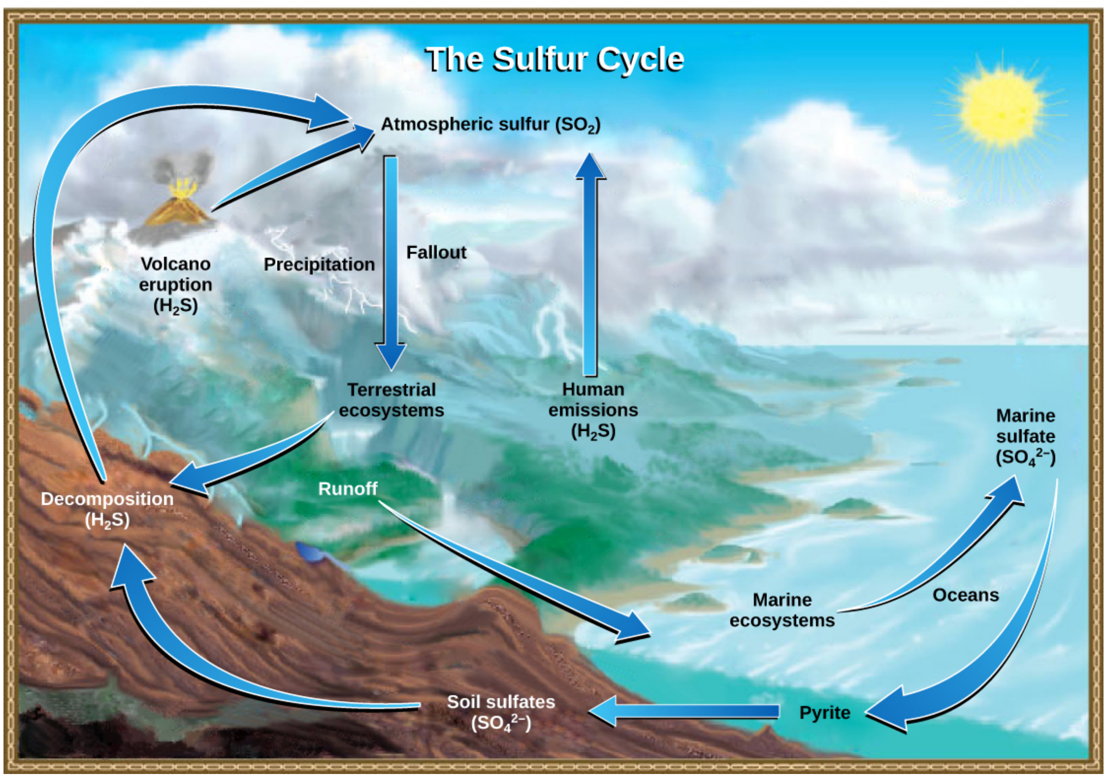
The restoration of the oyster population in the Chesapeake Bay has been ongoing for several years with mixed success. Not only do many people find oysters good to eat, but the oysters also clean up the bay. They are filter feeders, and as they eat, they clean the water around them. Filter feeders eat by pumping a continuous stream of water over finely divided appendages (gills in the case of oysters) and capturing prokaryotes, plankton, and fine organic particles in their mucus. In the 1700s, it was estimated that it took only a few days for the oyster population to filter the entire volume of the bay. Today, with the changed water conditions, it is estimated that the present population would take nearly a year to do the same job.

Restoration efforts have been ongoing for several years by non-profit organizations such as the Chesapeake Bay Foundation. The restoration goal is to find a way to increase population density so the oysters can reproduce more efficiently. Many disease-resistant varieties (developed at the Virginia Institute of Marine Science for the College of William and Mary) are now available and have been used in the construction of experimental oyster reefs. Efforts by Virginia and Delaware to clean and restore the bay have been hampered because much of the pollution entering the bay comes from other states, which emphasizes the need for interstate cooperation to gain successful restoration.

The new, hearty oyster strains have also spawned a new and economically viable industry—oyster aquaculture—which not only supplies oysters for food and profit, but also has the added benefit of cleaning the bay.

The Sulfur Cycle

Sulfur is an essential element for the macromolecules of living things. As part of the amino acid cysteine, it is involved in the formation of proteins. As shown in [link], sulfur cycles between the oceans, land, and atmosphere. Atmospheric sulfur is found in the form of sulfur dioxide (SO_2), which enters the atmosphere in three ways: first, from the decomposition of organic molecules; second, from volcanic activity and geothermal vents; and, third, from the burning of fossil fuels by humans.



Sulfur dioxide from the atmosphere becomes available to terrestrial and marine ecosystems when it is dissolved in

precipitation as weak sulfuric acid or when it falls directly to Earth as fallout. Weathering of rocks also makes sulfates available to terrestrial ecosystems. Decomposition of living organisms returns sulfates to the ocean, soil, and atmosphere. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

On land, sulfur is deposited in four major ways: precipitation, direct fallout from the atmosphere, rock weathering, and geothermal vents ([link]). Atmospheric sulfur is found in the form of sulfur dioxide (SO_2), and as rain falls through the atmosphere, sulfur is dissolved in the form of weak sulfuric acid (H_2SO_4). Sulfur can also fall directly from the atmosphere in a process called **fallout**. Also, as sulfur-containing rocks weather, sulfur is released into the soil. These rocks originate from ocean sediments that are moved to land by the geologic uplifting of ocean sediments. Terrestrial ecosystems can then make use of these soil sulfates (SO_4^{2-}), which enter the food web by being taken up by plant roots. When these plants decompose and die, sulfur is released back into the atmosphere as hydrogen sulfide (H_2S) gas.

At this sulfur vent in Lassen Volcanic

National Park in northeastern California, the yellowish sulfur deposits are visible near the mouth of the vent. (credit: “Calbear22”/Wikimedia Commons)

Sulfur enters the ocean in runoff from land, from atmospheric fallout, and from underwater geothermal vents. Some ecosystems rely on chemoautotrophs using sulfur as a biological energy source. This sulfur then supports marine ecosystems in the form of sulfates.

Human activities have played a major role in altering the balance of the global sulfur cycle. The burning of large quantities of fossil fuels, especially from coal, releases larger amounts of hydrogen sulfide gas into the atmosphere. As rain falls through this gas, it creates the phenomenon known as acid rain, which damages the natural environment by lowering the pH of lakes, thus killing many of the resident plants and animals. **Acid rain** is corrosive rain caused by rainwater falling to the ground through sulfur dioxide gas, turning it into weak sulfuric acid, which causes damage to aquatic ecosystems. Acid rain also affects the man-made environment through the chemical degradation of buildings. For example, many marble monuments, such as the Lincoln Memorial in Washington, DC, have suffered significant damage from acid rain over the years. These examples show the wide-ranging effects of human activities on our environment and the challenges that remain for our future.

Section Summary

Mineral nutrients are cycled through ecosystems and their environment. Of particular importance are water, carbon, nitrogen, phosphorus, and sulfur. All of these cycles have major impacts on ecosystem structure and function. As human activities have caused major disturbances to these cycles, their study and modeling is especially important. Ecosystems have been damaged by a variety of human activities that alter the natural biogeochemical cycles due to pollution, oil spills, and events causing global

climate change. The health of the biosphere depends on understanding these cycles and how to protect the environment from irreversible damage.

Art Connections

Exercise:

Problem:

[link] Which of the following statements about the nitrogen cycle is false?

a. Ammonification converts organic nitrogenous matter from living organisms into ammonium (NH_4^+).
b. Denitrification by bacteria converts nitrates (NO_3^-) to nitrogen gas (N_2).
c. Nitrification by bacteria converts nitrates (NO_3^-) to nitrites (NO_2^-).
d. Nitrogen fixing bacteria convert nitrogen gas (N_2) into organic compounds.

Solution:

[link] C: Nitrification by bacteria converts nitrates (NO_3^-) to nitrites (NO_2^-).

Multiple Choice

Exercise:

Problem: The majority of the water found on Earth is:

a. ice
b. water vapor
c. fresh water
d. salt water

Solution:

D

Exercise:

Problem:

The process whereby oxygen is depleted by the growth of microorganisms due to excess nutrients in aquatic systems is called ________.

a. dead zoning
b. eutrophication
c. retrophication
d. depletion

Solution:

B

Free Response

Exercise:

Problem:

Why are drinking water supplies still a major concern for many countries?

Solution:

Most of the water on Earth is salt water, which humans cannot drink unless the salt is removed. Some fresh water is locked in glaciers and polar ice caps, or is present in the atmosphere. The earth’s water supplies are threatened by pollution and exhaustion. The effort to

supply fresh drinking water to the planet's ever-expanding human population is seen as a major challenge in this century.

Glossary

acid rain
: a corrosive rain caused by rainwater mixing with sulfur dioxide gas as it fall through the atmosphere, turning it into weak sulfuric acid, causing damage to aquatic ecosystems

biogeochemical cycle
: the cycling of minerals and nutrients through the biotic and abiotic world

dead zone
: an area in a lake and ocean near the mouths of rivers where large areas are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities

eutrophication
: the process whereby nutrient runoff causes the excess growth of microorganisms and plants in aquatic systems

fallout
: the direct deposition of solid minerals on land or in the ocean from the atmosphere

hydrosphere
: the region of the planet in which water exists, including the atmosphere that contains water vapor and the region beneath the ground that contains groundwater

non-renewable resource
: a resource, such as a fossil fuel, that is either regenerated very slowly or not at all

subduction

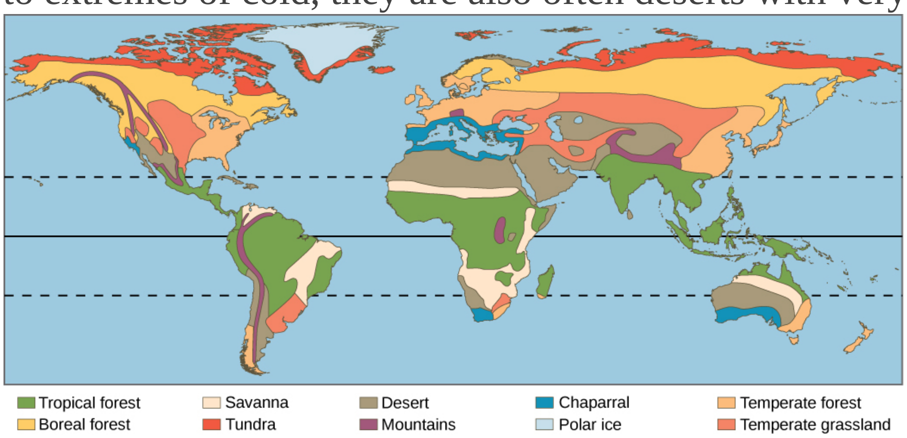
the movement of one tectonic plate beneath another

Terrestrial Biomes

By the end of this section, you will be able to:

- Identify the two major abiotic factors that determine the type of terrestrial biome in an area
- Recognize distinguishing characteristics of each of the eight major terrestrial biomes

Earth’s biomes can be either terrestrial or aquatic. Terrestrial biomes are based on land, while aquatic biomes include both ocean and freshwater biomes. The eight major terrestrial biomes on Earth are each distinguished by characteristic temperatures and amount of precipitation. Annual totals and fluctuations of precipitation affect the kinds of vegetation and animal life that can exist in broad geographical regions. Temperature variation on a daily and seasonal basis is also important for predicting the geographic distribution of a biome. Since a biome is defined by climate, the same biome can occur in geographically distinct areas with similar climates ([link]). There are also large areas on Antarctica, Greenland, and in mountain ranges that are covered by permanent glaciers and support very little life. Strictly speaking, these are not considered biomes and in addition to extremes of cold, they are also often deserts with very low precipitation.



Each of the world’s eight major biomes is distinguished by characteristic temperatures and amount of precipitation. Polar ice caps and mountains are also shown.

Tropical Forest

Tropical rainforests are also referred to as tropical wet forests. This biome is found in equatorial regions ([link]). Tropical rainforests are the most diverse terrestrial biome. This biodiversity is still largely unknown to science and is under extraordinary threat primarily through logging and deforestation for agriculture. Tropical rainforests have also been described as nature's pharmacy because of the potential for new drugs that is largely hidden in the chemicals produced by the huge diversity of plants, animals, and other organisms. The vegetation is characterized by plants with spreading roots and broad leaves that fall off throughout the year, unlike the trees of deciduous forests that lose their leaves in one season. These forests are "evergreen," year-round.

The temperature and sunlight profiles of tropical rainforests are stable in comparison to that of other terrestrial biomes, with average temperatures ranging from 20^{o}C to 34^{o}C (68^{o}F to 93^{o}F). Month-to-month temperatures are relatively constant in tropical rainforests, in contrast to forests further from the equator. This lack of temperature seasonality leads to year-round plant growth, rather than the seasonal growth seen in other biomes. In contrast to other ecosystems, a more constant daily amount of sunlight (11–12 hours per day) provides more solar radiation, thereby a longer period of time for plant growth.

The annual rainfall in tropical rainforests ranges from 250 cm to more than 450 cm (8.2–14.8 ft) with considerable seasonal variation. Tropical rainforests have wet months in which there can be more than 30 cm (11–12 in) of precipitation, as well as dry months in which there are fewer than 10 cm (3.5 in) of rainfall. However, the driest month of a tropical rainforest can still exceed the *annual* rainfall of some other biomes, such as deserts.

Tropical rainforests have high net primary productivity because the annual temperatures and precipitation values support rapid plant growth ([link]). However, the high rainfall quickly leaches nutrients from the soils of these forests, which are typically low in nutrients. Tropical rainforests are characterized by vertical layering of vegetation and the formation of distinct habitats for animals within each layer. On the forest floor is a sparse layer

of plants and decaying plant matter. Above that is an understory of short, shrubby foliage. A layer of trees rises above this understory and is topped by a closed upper **canopy**—the uppermost overhead layer of branches and leaves. Some additional trees emerge through this closed upper canopy. These layers provide diverse and complex habitats for the variety of plants, animals, and other organisms within the tropical wet forests. Many species of animals use the variety of plants and the complex structure of the tropical wet forests for food and shelter. Some organisms live several meters above ground rarely ever descending to the forest floor.

Rainforests are not the only forest biome in the tropics; there are also tropical dry forests, which are characterized by a dry season of varying lengths. These forests commonly experience leaf loss during the dry season to one degree or another. The loss of leaves from taller trees during the dry season opens up the canopy and allows sunlight to the forest floor that allows the growth of thick ground-level brush, which is absent in tropical rainforests. Extensive tropical dry forests occur in Africa (including Madagascar), India, southern Mexico, and South America.

Species diversity is very high in tropical wet forests, such as these forests of Madre de

Dios, Peru, near the Amazon River. (credit: Roosevelt Garcia)

Savannas

Savannas are grasslands with scattered trees, and they are found in Africa, South America, and northern Australia ([link]). Savannas are hot, tropical areas with temperatures averaging from 24°C –29°C (75°F –84°F) and an annual rainfall of 51–127 cm (20–50 in). Savannas have an extensive dry season and consequent fires. As a result, scattered in the grasses and forbs (herbaceous flowering plants) that dominate the savanna, there are relatively few trees ([link]). Since fire is an important source of disturbance in this biome, plants have evolved well-developed root systems that allow them to quickly re-sprout after a fire.

Although savannas are dominated by grasses, small woodlands, such as this one in Mount Archer National Park in Queensland,

Australia, may dot the landscape. (credit: "Ethel Aardvark"/Wikimedia Commons)

Deserts

Subtropical deserts exist between 15^o and 30^o north and south latitude and are centered on the Tropic of Cancer and the Tropic of Capricorn ([link]). Deserts are frequently located on the downwind or lee side of mountain ranges, which create a rain shadow after prevailing winds drop their water content on the mountains. This is typical of the North American deserts, such as the Mohave and Sonoran deserts. Deserts in other regions, such as the Sahara Desert in northern Africa or the Namib Desert in southwestern Africa are dry because of the high-pressure, dry air descending at those latitudes. Subtropical deserts are very dry; evaporation typically exceeds precipitation. Subtropical hot deserts can have daytime soil surface temperatures above 60^oC (140^oF) and nighttime temperatures approaching 0^oC (32^oF). The temperature drops so far because there is little water vapor in the air to prevent radiative cooling of the land surface. Subtropical deserts are characterized by low annual precipitation of fewer than 30 cm (12 in) with little monthly variation and lack of predictability in rainfall. Some years may receive tiny amounts of rainfall, while others receive more. In some cases, the annual rainfall can be as low as 2 cm (0.8 in) in subtropical deserts located in central Australia (“the Outback”) and northern Africa.

The low species diversity of this biome is closely related to its low and unpredictable precipitation. Despite the relatively low diversity, desert species exhibit fascinating adaptations to the harshness of their environment. Very dry deserts lack perennial vegetation that lives from one year to the next; instead, many plants are annuals that grow quickly and reproduce when rainfall does occur, then they die. Perennial plants in deserts are characterized by adaptations that conserve water: deep roots, reduced foliage, and water-storing stems ([link]). Seed plants in the desert produce seeds that can lie dormant for extended periods between rains. Most animal life in subtropical deserts has adapted to a nocturnal life, spending the hot daytime hours beneath the ground. The Namib Desert is

the oldest on the planet, and has probably been dry for more than 55 million years. It supports a number of endemic species (species found only there) because of this great age. For example, the unusual gymnosperm *Welwitschia mirabilis* is the only extant species of an entire order of plants. There are also five species of reptiles considered endemic to the Namib.

In addition to subtropical deserts there are cold deserts that experience freezing temperatures during the winter and any precipitation is in the form of snowfall. The largest of these deserts are the Gobi Desert in northern China and southern Mongolia, the Taklimakan Desert in western China, the Turkestan Desert, and the Great Basin Desert of the United States.

Many desert plants have tiny leaves or no leaves at all to reduce water loss. The leaves of ocotillo, shown here in the Chihuahuan Desert in Big Bend National Park, Texas, appear only after rainfall and then are shed. (credit “bare ocotillo”: "Leaflet"/Wikimedia Commons)

Chaparral

The **chaparral** is also called scrub forest and is found in California, along the Mediterranean Sea, and along the southern coast of Australia ([link]). The annual rainfall in this biome ranges from 65 cm to 75 cm (25.6–29.5

in) and the majority of the rain falls in the winter. Summers are very dry and many chaparral plants are dormant during the summertime. The chaparral vegetation is dominated by shrubs and is adapted to periodic fires, with some plants producing seeds that germinate only after a hot fire. The ashes left behind after a fire are rich in nutrients like nitrogen that fertilize the soil and promote plant regrowth. Fire is a natural part of the maintenance of this biome and frequently threatens human habitation in this biome in the U.S. ([link]).

The chaparral is dominated by shrubs. (credit: Miguel Vieira)

Temperate Grasslands

Temperate grasslands are found throughout central North America, where they are also known as prairies, and in Eurasia, where they are known as steppes ([link]). Temperate grasslands have pronounced annual fluctuations in temperature with hot summers and cold winters. The annual temperature variation produces specific growing seasons for plants. Plant growth is

possible when temperatures are warm enough to sustain plant growth, which occurs in the spring, summer, and fall.

Annual precipitation ranges from 25.4 cm to 88.9 cm (10–35 in). Temperate grasslands have few trees except for those found growing along rivers or streams. The dominant vegetation tends to consist of grasses. The treeless condition is maintained by low precipitation, frequent fires, and grazing ([link]). The vegetation is very dense and the soils are fertile because the subsurface of the soil is packed with the roots and rhizomes (underground stems) of these grasses. The roots and rhizomes act to anchor plants into the ground and replenish the organic material (humus) in the soil when they die and decay.

The American bison (*Bison bison*), more commonly called the buffalo, is a grazing mammal that once populated American prairies in huge numbers. (credit: Jack Dykinga, USDA ARS)

Fires, which are a natural disturbance in temperate grasslands, can be ignited by lightning strikes. It also appears that the lightning-caused fire

regime in North American grasslands was enhanced by intentional burning by humans. When fire is suppressed in temperate grasslands, the vegetation eventually converts to scrub and dense forests. Often, the restoration or management of temperate grasslands requires the use of controlled burns to suppress the growth of trees and maintain the grasses.

Temperate Forests

Temperate forests are the most common biome in eastern North America, Western Europe, Eastern Asia, Chile, and New Zealand ([link]). This biome is found throughout mid-latitude regions. Temperatures range between –30°C and 30°C (–22°F to 86°F) and drop to below freezing on an annual basis. These temperatures mean that temperate forests have defined growing seasons during the spring, summer, and early fall. Precipitation is relatively constant throughout the year and ranges between 75 cm and 150 cm (29.5–59 in).

Deciduous trees are the dominant plant in this biome with fewer evergreen conifers. Deciduous trees lose their leaves each fall and remain leafless in the winter. Thus, little photosynthesis occurs during the dormant winter period. Each spring, new leaves appear as temperature increases. Because of the dormant period, the net primary productivity of temperate forests is less than that of tropical rainforests. In addition, temperate forests show far less diversity of tree species than tropical rainforest biomes.

The trees of the temperate forests leaf out and shade much of the ground; however, more sunlight reaches the ground in this biome than in tropical rainforests because trees in temperate forests do not grow as tall as the trees in tropical rainforests. The soils of the temperate forests are rich in inorganic and organic nutrients compared to tropical rainforests. This is because of the thick layer of leaf litter on forest floors and reduced leaching of nutrients by rainfall. As this leaf litter decays, nutrients are returned to the soil. The leaf litter also protects soil from erosion, insulates the ground, and provides habitats for invertebrates and their predators ([link]).

Deciduous trees are the dominant plant in the temperate forest. (credit: Oliver Herold)

Boreal Forests

The **boreal forest**, also known as taiga or coniferous forest, is found roughly between 50° and 60° north latitude across most of Canada, Alaska, Russia, and northern Europe ([link]). Boreal forests are also found above a certain elevation (and below high elevations where trees cannot grow) in mountain ranges throughout the Northern Hemisphere. This biome has cold, dry winters and short, cool, wet summers. The annual precipitation is from 40 cm to 100 cm (15.7–39 in) and usually takes the form of snow; little evaporation occurs because of the cold temperatures.

The long and cold winters in the boreal forest have led to the predominance of cold-tolerant cone-bearing plants. These are evergreen coniferous trees like pines, spruce, and fir, which retain their needle-shaped leaves year-round. Evergreen trees can photosynthesize earlier in the spring than deciduous trees because less energy from the Sun is required to warm a needle-like leaf than a broad leaf. Evergreen trees grow faster than

deciduous trees in the boreal forest. In addition, soils in boreal forest regions tend to be acidic with little available nitrogen. Leaves are a nitrogen-rich structure and deciduous trees must produce a new set of these nitrogen-rich structures each year. Therefore, coniferous trees that retain nitrogen-rich needles in a nitrogen limiting environment may have had a competitive advantage over the broad-leafed deciduous trees.

The net primary productivity of boreal forests is lower than that of temperate forests and tropical wet forests. The aboveground biomass of boreal forests is high because these slow-growing tree species are long-lived and accumulate standing biomass over time. Species diversity is less than that seen in temperate forests and tropical rainforests. Boreal forests lack the layered forest structure seen in tropical rainforests or, to a lesser degree, temperate forests. The structure of a boreal forest is often only a tree layer and a ground layer. When conifer needles are dropped, they decompose more slowly than broad leaves; therefore, fewer nutrients are returned to the soil to fuel plant growth ([link]).

The boreal forest (taiga) has low lying plants and conifer trees. (credit: L.B. Brubaker, NOAA)

Arctic Tundra

The **Arctic tundra** lies north of the subarctic boreal forests and is located throughout the Arctic regions of the Northern Hemisphere ([link]). Tundra also exists at elevations above the tree line on mountains. The average winter temperature is –34°C (–29.2°F) and the average summer temperature is 3°C–12°C (37°F –52°F). Plants in the Arctic tundra have a short growing season of approximately 50–60 days. However, during this time, there are almost 24 hours of daylight and plant growth is rapid. The annual precipitation of the Arctic tundra is low (15–25 cm or 6–10 in) with little annual variation in precipitation. And, as in the boreal forests, there is little evaporation because of the cold temperatures.

Plants in the Arctic tundra are generally low to the ground and include low shrubs, grasses, lichens, and small flowering plants ([link]). There is little species diversity, low net primary productivity, and low aboveground biomass. The soils of the Arctic tundra may remain in a perennially frozen state referred to as **permafrost**. The permafrost makes it impossible for roots to penetrate far into the soil and slows the decay of organic matter, which inhibits the release of nutrients from organic matter. The melting of the permafrost in the brief summer provides water for a burst of productivity while temperatures and long days permit it. During the growing season, the ground of the Arctic tundra can be completely covered with plants or lichens.

Low-growing plants such as shrub willow dominate the tundra landscape during the summer, shown here in the Arctic National Wildlife Refuge. (credit: Arctic National Wildlife Refuge, USFWS)

Note:
Concept in Action

Watch this *Assignment Discovery: Biomes* video for an overview of biomes. To explore further, select one of the biomes on the extended playlist: desert, savanna, temperate forest, temperate grassland, tropic, tundra.

Section Summary

Earth has terrestrial and aquatic biomes. Aquatic biomes include both freshwater and marine environments. There are eight major terrestrial biomes: tropical rainforests, savannas, subtropical deserts, chaparral, temperate grasslands, temperate forests, boreal forests, and Arctic tundra. The same biome can occur in different geographic locations with similar climates. Temperature and precipitation, and variations in both, are key abiotic factors that shape the composition of animal and plant communities in terrestrial biomes. Some biomes, such as temperate grasslands and temperate forests, have distinct seasons with cold and hot weather alternating throughout the year. In warm, moist biomes, such as the tropical rainforest, net primary productivity is high as warm temperatures, abundant water, and a year-round growing season fuel plant growth. Other biomes, such as deserts and tundra, have low primary productivity due to extreme temperatures and a shortage of water.

Multiple Choice

Exercise:

Problem:

Which of the following biomes is characterized by abundant water resources?

a. deserts
b. boreal forests
c. savanna
d. tropical wet forests

Solution:

D

Exercise:

Problem:

Which of the following biomes is characterized by short growing seasons?

a. deserts
b. tropical wet forests
c. Arctic tundra
d. savanna

Solution:

C

Exercise:

Problem: Why is the tundra treeless?

a. lack of sufficient water
b. permanently frozen ground
c. winters too harsh
d. too many fires

Solution:

B

Free Response

Exercise:

Problem:

The extremely low precipitation of subtropical desert biomes might lead one to expect fire to be a major disturbance factor; however, fire is more common in the temperate grassland biome than in the subtropical desert biome. Why is this?

Solution:

Fire is less common in desert biomes than in temperate grasslands because deserts have low net primary productivity, thus very little plant biomass to fuel a fire.

Exercise:

Problem:

In what ways are the subtropical desert and the Arctic tundra similar?

Solution:

Both the subtropical desert and the Arctic tundra have a low supply of water. In the desert, this is due to extremely low precipitation, and in the Arctic tundra, much of the water is unavailable to plants because it is frozen. Both the subtropical desert and the Arctic tundra have low net primary productivity.

Glossary

arctic tundra
: a biome characterized by low average temperatures, brief growing seasons, the presence of permafrost, and limited precipitation largely in the form of snow in which the dominant vegetation are low shrubs, lichens, mosses, and small herbaceous plants

boreal forest
: a biome found in temperate and subarctic regions characterized by short growing seasons and dominated structurally by coniferous trees

canopy
: the branches and foliage of trees that form a layer of overhead coverage in a forest

chaparral
: a biome found in temperate coastal regions characterized by low trees and dry-adapted shrubs and forbs

permafrost
: a perennially frozen portion of the Arctic tundra soil

savanna
: a biome located in the tropics with an extended dry season and characterized by a grassland with sparsely distributed trees

subtropical desert
: a biome found in the subtropics with hot daily temperatures, very low and unpredictable precipitation, and characterized by a limited dry-adapted vegetation

temperate forest
: a biome found in temperate regions with moderate rainfall and dominated structurally by deciduous trees

temperate grassland
: a biome dominated by grasses and herbaceous plants due to low precipitation, periodic fires, and grazing

tropical rainforest
: a biome found near the equator characterized by stable temperatures with abundant and seasonal rainfall in which trees form the structurally important vegetation

Aquatic and Marine Biomes

By the end of this section, you will be able to:

- Describe the effects of abiotic factors on the composition of plant and animal communities in aquatic biomes
- Compare the characteristics of the ocean zones
- Summarize the characteristics of standing water and flowing water in freshwater biomes

Like terrestrial biomes, aquatic biomes are influenced by abiotic factors. In the case of aquatic biomes the abiotic factors include light, temperature, flow regime, and dissolved solids. The aquatic medium—water— has different physical and chemical properties than air. Even if the water in a pond or other body of water is perfectly clear (there are no suspended particles), water, on its own, absorbs light. As one descends deep enough into a body of water, eventually there will be a depth at which the sunlight cannot reach. While there are some abiotic and biotic factors in a terrestrial ecosystem that shade light (like fog, dust, or insect swarms), these are not usually permanent features of the environment. The importance of light in aquatic biomes is central to the communities of organisms found in both freshwater and marine ecosystems because it controls productivity through photosynthesis.

In addition to light, solar radiation warms bodies of water and many exhibit distinct layers of water at differing temperatures. The water temperature affects the organisms' rates of growth and the amount of dissolved oxygen available for respiration.

The movement of water is also important in many aquatic biomes. In rivers, the organisms must obviously be adapted to the constant movement of the water around them, but even in larger bodies of water such as the oceans, regular currents and tides impact availability of nutrients, food resources, and the presence of the water itself.

Finally, all natural water contains dissolved solids, or salts. Fresh water contains low levels of such dissolved substances because the water is rapidly recycled through evaporation and precipitation. The oceans have a relatively constant high salt content. Aquatic habitats at the interface of

marine and freshwater ecosystems have complex and variable salt environments that range between freshwater and marine levels. These are known as brackish water environments. Lakes located in closed drainage basins concentrate salt in their waters and can have extremely high salt content that only a few and highly specialized species are able to inhabit.

Marine Biomes

The ocean is a continuous body of salt water that is relatively uniform in chemical composition. It is a weak solution of mineral salts and decayed biological matter. Within the ocean, coral reefs are a second type of marine biome. Estuaries, coastal areas where salt water and fresh water mix, form a third unique marine biome.

The ocean is categorized by several zones ([link]). All of the ocean's open water is referred to as the **pelagic realm** (or zone). The **benthic realm** (or zone) extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor. From the surface to the bottom or the limit to which photosynthesis occurs is the **photic zone** (approximately 200 m or 650 ft). At depths greater than 200 m, light cannot penetrate; thus, this is referred to as the **aphotic zone**. The majority of the ocean is aphotic and lacks sufficient light for photosynthesis. The deepest part of the ocean, the Challenger Deep (in the Mariana Trench, located in the western Pacific Ocean), is about 11,000 m (about 6.8 mi) deep. To give some perspective on the depth of this trench, the ocean is, on average, 4267 m or 14,000 ft deep.

Ocean

The physical diversity of the ocean has a significant influence on the diversity of organisms that live within it. The ocean is categorized into different zones based on how far light reaches into the water. Each zone has a distinct group of species adapted to the biotic and abiotic conditions particular to that zone.

The **intertidal zone** ([link]) is the oceanic region that is closest to land. With each tidal cycle, the intertidal zone alternates between being inundated

with water and left high and dry. Generally, most people think of this portion of the ocean as a sandy beach. In some cases, the intertidal zone is indeed a sandy beach, but it can also be rocky, muddy, or dense with tangled roots in mangrove forests. The intertidal zone is an extremely variable environment because of tides. Organisms may be exposed to air at low tide and are underwater during high tide. Therefore, living things that thrive in the intertidal zone are often adapted to being dry for long periods of time. The shore of the intertidal zone is also repeatedly struck by waves and the organisms found there are adapted to withstand damage from the pounding action of the waves ([link]). The exoskeletons of shoreline crustaceans (such as the shore crab, *Carcinus maenas*) are tough and protect them from desiccation (drying out) and wave damage. Another consequence of the pounding waves is that few algae and plants establish themselves in constantly moving sand or mud.

Sea stars, sea urchins, and mussel shells are often found in the intertidal zone, shown here in Kachemak Bay, Alaska. (credit: NOAA)

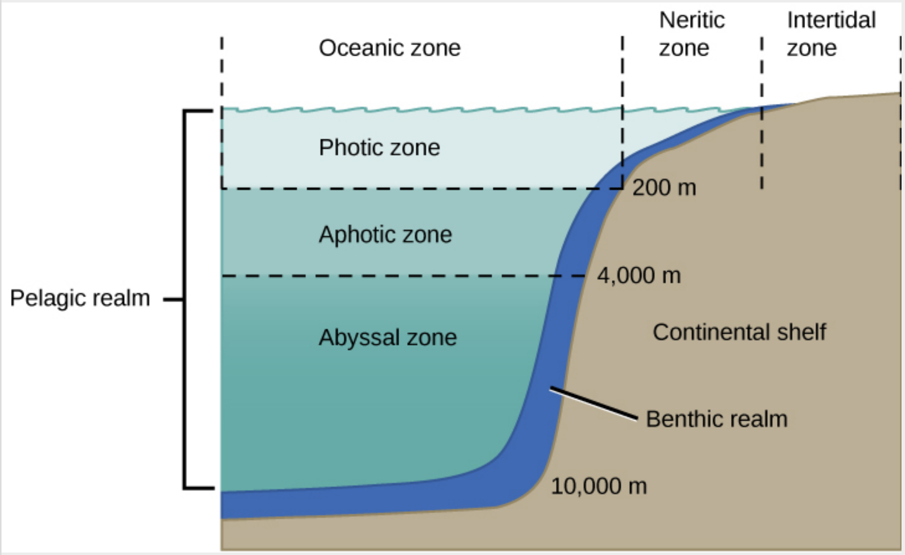
The **neritic zone** ([link]) extends from the margin of the intertidal zone to depths of about 200 m (or 650 ft) at the edge of the continental shelf. When the water is relatively clear, photosynthesis can occur in the neritic zone. The water contains silt and is well-oxygenated, low in pressure, and stable in temperature. These factors all contribute to the neritic zone having the highest productivity and biodiversity of the ocean. Phytoplankton, including photosynthetic bacteria and larger species of algae, are responsible for the bulk of this primary productivity. Zooplankton, protists, small fishes, and shrimp feed on the producers and are the primary food source for most of the world’s fisheries. The majority of these fisheries exist within the neritic zone.

Beyond the neritic zone is the open ocean area known as the **oceanic zone** ([link]). Within the oceanic zone there is thermal stratification. Abundant phytoplankton and zooplankton support populations of fish and whales. Nutrients are scarce and this is a relatively less productive part of the marine biome. When photosynthetic organisms and the organisms that feed on them die, their bodies fall to the bottom of the ocean where they remain; the open ocean lacks a process for bringing the organic nutrients back up to the surface.

Beneath the pelagic zone is the benthic realm, the deepwater region beyond the continental shelf ([link]). The bottom of the benthic realm is comprised of sand, silt, and dead organisms. Temperature decreases as water depth increases. This is a nutrient-rich portion of the ocean because of the dead organisms that fall from the upper layers of the ocean. Because of this high level of nutrients, a diversity of fungi, sponges, sea anemones, marine worms, sea stars, fishes, and bacteria exists.

The deepest part of the ocean is the **abyssal zone**, which is at depths of 4000 m or greater. The abyssal zone ([link]) is very cold and has very high pressure, high oxygen content, and low nutrient content. There are a variety of invertebrates and fishes found in this zone, but the abyssal zone does not have photosynthetic organisms. Chemosynthetic bacteria use the hydrogen sulfide and other minerals emitted from deep hydrothermal vents. These chemosynthetic bacteria use the hydrogen sulfide as an energy source and serve as the base of the food chain found around the vents.

Note:
Art Connection



The ocean is divided into different zones based on water depth, distance from the shoreline, and light penetration.

In which of the following regions would you expect to find photosynthetic organisms?

a. The aphotic zone, the neritic zone, the oceanic zone, and the benthic realm.
b. The photic zone, the intertidal zone, the neritic zone, and the oceanic zone.
c. The photic zone, the abyssal zone, the neritic zone, and the oceanic zone.
d. The pelagic realm, the aphotic zone, the neritic zone, and the oceanic zone.

Coral Reefs

Coral reefs are ocean ridges formed by marine invertebrates living in warm shallow waters within the photic zone of the ocean. They are found within 30˚ north and south of the equator. The Great Barrier Reef is a well-known reef system located several miles off the northeastern coast of Australia. Other coral reefs are fringing islands, which are directly adjacent to land, or atolls, which are circular reefs surrounding a former island that is now underwater. The coral-forming colonies of organisms (members of phylum Cnidaria) secrete a calcium carbonate skeleton. These calcium-rich skeletons slowly accumulate, thus forming the underwater reef ([link]). Corals found in shallower waters (at a depth of approximately 60 m or about 200 ft) have a mutualistic relationship with photosynthetic unicellular protists. The relationship provides corals with the majority of the nutrition and the energy they require. The waters in which these corals live are nutritionally poor and, without this mutualism, it would not be possible for large corals to grow because there are few planktonic organisms for them to feed on. Some corals living in deeper and colder water do not have a mutualistic relationship with protists; these corals must obtain their energy exclusively by feeding on plankton using stinging cells on their tentacles.

Note:
Concept in Action

In this National Oceanic and Atmospheric Administration (NOAA) video, marine ecologist Dr. Peter Etnoyer discusses his research on coral organisms.

Coral reefs are one of the most diverse biomes. It is estimated that more than 4000 fish species inhabit coral reefs. These fishes can feed on coral,

the **cryptofauna** (invertebrates found within the calcium carbonate structures of the coral reefs), or the seaweed and algae that are associated with the coral. These species include predators, herbivores, or planktivores. Predators are animal species that hunt and are carnivores or “flesh eaters.” Herbivores eat plant material, and **planktivores** eat plankton.

Coral reefs are formed by the calcium carbonate skeletons of coral organisms, which are marine invertebrates in the phylum Cnidaria. (credit: Terry Hughes)

Note:

Evolution in Action

Global Decline of Coral Reefs

It takes a long time to build a coral reef. The animals that create coral reefs do so over thousands of years, continuing to slowly deposit the calcium carbonate that forms their characteristic ocean homes. Bathed in warm tropical waters, the coral animals and their symbiotic protist partners evolved to survive at the upper limit of ocean water temperature.

Together, climate change and human activity pose dual threats to the long-term survival of the world’s coral reefs. The main cause of killing of coral reefs is warmer-than-usual surface water. As global warming raises ocean temperatures, coral reefs are suffering. The excessive warmth causes the coral organisms to expel their endosymbiotic, food-producing protists, resulting in a phenomenon known as bleaching. The colors of corals are a result of the particular protist endosymbiont, and when the protists leave, the corals lose their color and turn white, hence the term “bleaching.” Rising levels of atmospheric carbon dioxide further threaten the corals in other ways; as carbon dioxide dissolves in ocean waters, it lowers pH, thus increasing ocean acidity. As acidity increases, it interferes with the calcification that normally occurs as coral animals build their calcium carbonate homes.

When a coral reef begins to die, species diversity plummets as animals lose food and shelter. Coral reefs are also economically important tourist destinations, so the decline of coral reefs poses a serious threat to coastal economies.

Human population growth has damaged corals in other ways, too. As human coastal populations increase, the runoff of sediment and agricultural chemicals has increased, causing some of the once-clear tropical waters to become cloudy. At the same time, overfishing of popular fish species has allowed the predator species that eat corals to go unchecked.

Although a rise in global temperatures of 1°C–2°C (a conservative scientific projection) in the coming decades may not seem large, it is very significant to this biome. When change occurs rapidly, species can become extinct before evolution leads to newly adapted species. Many scientists believe that global warming, with its rapid (in terms of evolutionary time) and inexorable increases in temperature, is tipping the balance beyond the point at which many of the world’s coral reefs can recover.

Estuaries: Where the Ocean Meets Fresh Water

Estuaries are biomes that occur where a river, a source of fresh water, meets the ocean. Therefore, both fresh water and salt water are found in the same vicinity; mixing results in a diluted (brackish) salt water. Estuaries form protected areas where many of the offspring of crustaceans, mollusks, and fish begin their lives. Salinity is an important factor that influences the organisms and the adaptations of the organisms found in estuaries. The salinity of estuaries varies and is based on the rate of flow of its freshwater sources. Once or twice a day, high tides bring salt water into the estuary. Low tides occurring at the same frequency reverse the current of salt water ([link]).

As estuary is where fresh water and salt water meet, such as the mouth of the Klamath River in California, shown here. (credit: U.S. Army Corps of Engineers)

The daily mixing of fresh water and salt water is a physiological challenge for the plants and animals that inhabit estuaries. Many estuarine plant species are halophytes, plants that can tolerate salty conditions. Halophytic

plants are adapted to deal with salt water spray and salt water on their roots. In some halophytes, filters in the roots remove the salt from the water that the plant absorbs. Animals, such as mussels and clams (phylum Mollusca), have developed behavioral adaptations that expend a lot of energy to function in this rapidly changing environment. When these animals are exposed to low salinity, they stop feeding, close their shells, and switch from aerobic respiration (in which they use gills) to anaerobic respiration (a process that does not require oxygen). When high tide returns to the estuary, the salinity and oxygen content of the water increases, and these animals open their shells, begin feeding, and return to aerobic respiration.

Freshwater Biomes

Freshwater biomes include lakes, ponds, and wetlands (standing water) as well as rivers and streams (flowing water). Humans rely on freshwater biomes to provide aquatic resources for drinking water, crop irrigation, sanitation, recreation, and industry. These various roles and human benefits are referred to as **ecosystem services**. Lakes and ponds are found in terrestrial landscapes and are therefore connected with abiotic and biotic factors influencing these terrestrial biomes.

Lakes and Ponds

Lakes and ponds can range in area from a few square meters to thousands of square kilometers. Temperature is an important abiotic factor affecting living things found in lakes and ponds. During the summer in temperate regions, thermal stratification of deep lakes occurs when the upper layer of water is warmed by the Sun and does not mix with deeper, cooler water. The process produces a sharp transition between the warm water above and cold water beneath. The two layers do not mix until cooling temperatures and winds break down the stratification and the water in the lake mixes from top to bottom. During the period of stratification, most of the productivity occurs in the warm, well-illuminated, upper layer, while dead organisms slowly rain down into the cold, dark layer below where decomposing bacteria and cold-adapted species such as lake trout exist.

Like the ocean, lakes and ponds have a photic layer in which photosynthesis can occur. Phytoplankton (algae and cyanobacteria) are found here and provide the base of the food web of lakes and ponds. Zooplankton, such as rotifers and small crustaceans, consume these phytoplankton. At the bottom of lakes and ponds, bacteria in the aphotic zone break down dead organisms that sink to the bottom.

Nitrogen and particularly phosphorus are important limiting nutrients in lakes and ponds. Therefore, they are determining factors in the amount of phytoplankton growth in lakes and ponds. When there is a large input of nitrogen and phosphorus (e.g., from sewage and runoff from fertilized lawns and farms), the growth of algae skyrockets, resulting in a large accumulation of algae called an **algal bloom**. Algal blooms ([link]) can become so extensive that they reduce light penetration in water. As a result, the lake or pond becomes aphotic and photosynthetic plants cannot survive. When the algae die and decompose, severe oxygen depletion of the water occurs. Fishes and other organisms that require oxygen are then more likely to die.

The uncontrolled growth

of algae in this waterway has resulted in an algal bloom.

Rivers and Streams

Rivers and the narrower streams that feed into the rivers are continuously moving bodies of water that carry water from the source or headwater to the mouth at a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America ([link]).

(a) (b)

Rivers range from (a) narrow and shallow to (b) wide and slow moving. (credit a: modification of work by Cory Zanker; credit b: modification of work by David DeHetre)

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as **source water**. The source water is usually cold, low in nutrients, and clear. The **channel** (the width of the river or stream) is narrower here than at any other place along

the length of the river or stream. Headwater streams are of necessity at a higher elevation than the mouth of the river and often originate in regions with steep grades leading to higher flow rates than lower elevation stretches of the river.

Faster-moving water and the short distance from its origin results in minimal silt levels in headwater streams; therefore, the water is clear. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. Photosynthesis may be further reduced by tree cover reaching over the narrow stream. This shading also keeps temperatures lower. An additional input of energy can come from leaves or other organic material that falls into a river or stream from the trees and other plants that border the water. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. The leaves also support a food chain of invertebrates that eat them and are in turn eaten by predatory invertebrates and fish. Plants and animals have adapted to this fast-moving water. For instance, leeches (phylum Annelida) have elongated bodies and suckers on both ends. These suckers attach to the substrate, keeping the leech anchored in place. In temperate regions, freshwater trout species (phylum Chordata) may be an important predator in these fast-moving and colder river and streams.

As the river or stream flows away from the source, the width of the channel gradually widens, the current slows, and the temperature characteristically increases. The increasing width results from the increased volume of water from more and more tributaries. Gradients are typically lower farther along the river, which accounts for the slowing flow. With increasing volume can come increased silt, and as the flow rate slows, the silt may settle, thus increasing the deposition of sediment. Phytoplankton can also be suspended in slow-moving water. Therefore, the water will not be as clear as it is near the source. The water is also warmer as a result of longer exposure to sunlight and the absence of tree cover over wider expanses between banks. Worms (phylum Annelida) and insects (phylum Arthropoda) can be found burrowing into the mud. Predatory vertebrates (phylum Chordata) include waterfowl, frogs, and fishes. In heavily silt-laden rivers, these predators must find food in the murky waters, and, unlike the trout in the clear waters at the source, these vertebrates cannot use vision as their primary sense to

find food. Instead, they are more likely to use taste or chemical cues to find prey.

When a river reaches the ocean or a large lake, the water typically slows dramatically and any silt in the river water will settle. Rivers with high silt content discharging into oceans with minimal currents and wave action will build deltas, low-elevation areas of sand and mud, as the silt settles onto the ocean bottom. Rivers with low silt content or in areas where ocean currents or wave action are high create estuarine areas where the fresh water and salt water mix.

Wetlands

Wetlands are environments in which the soil is either permanently or periodically saturated with water. Wetlands are different from lakes and ponds because wetlands exhibit a near continuous cover of emergent vegetation. **Emergent vegetation** consists of wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface. There are several types of wetlands including marshes, swamps, bogs, mudflats, and salt marshes ([link]).

Located in southern Florida, Everglades National Park is vast array of wetland environments, including sawgrass marshes, cypress swamps, and estuarine mangrove forests. Here, a great egret walks among cypress trees. (credit: NPS)

Freshwater marshes and swamps are characterized by slow and steady water flow. Bogs develop in depressions where water flow is low or nonexistent. Bogs usually occur in areas where there is a clay bottom with poor percolation. Percolation is the movement of water through the pores in the soil or rocks. The water found in a bog is stagnant and oxygen depleted because the oxygen that is used during the decomposition of organic matter is not replaced. As the oxygen in the water is depleted, decomposition slows. This leads to organic acids and other acids building up and lowering the pH of the water. At a lower pH, nitrogen becomes unavailable to plants. This creates a challenge for plants because nitrogen is an important limiting resource. Some types of bog plants (such as sundews, pitcher plants, and Venus flytraps) capture insects and extract the nitrogen from their bodies. Bogs have low net primary productivity because the water found in bogs has low levels of nitrogen and oxygen.

Section Summary

Aquatic biomes include both saltwater and freshwater biomes. The abiotic factors important for the structuring of aquatic biomes can be different than those seen in terrestrial biomes. Sunlight is an important factor in bodies of water, especially those that are very deep, because of the role of photosynthesis in sustaining certain organisms. Other important factors include temperature, water movement, and salt content. Oceans may be thought of as consisting of different zones based on water depth, distance from the shoreline, and light penetrance. Different kinds of organisms are adapted to the conditions found in each zone. Coral reefs are unique marine ecosystems that are home to a wide variety of species. Estuaries are found where rivers meet the ocean; their shallow waters provide nourishment and shelter for young crustaceans, mollusks, fishes, and many other species.

Freshwater biomes include lakes, ponds, rivers, streams, and wetlands. Bogs are an interesting type of wetland characterized by standing water, a lower pH, and a lack of nitrogen.

Art Connections

Exercise:

Problem:

[link] In which of the following regions would you expect to find photosynthetic organisms?

a. The aphotic zone, the neritic zone, the oceanic zone, and the benthic realm.
b. The photic zone, the intertidal zone, the neritic zone, and the oceanic zone.
c. The photic zone, the abyssal zone, the neritic zone, and the oceanic zone.
d. The pelagic realm, the aphotic zone, the neritic zone, and the oceanic zone.

Solution:

[link] B. The photic zone, the intertidal zone, the neritic zone, and the oceanic zone.

Multiple Choice

Exercise:

Problem:

Where would you expect to find the most photosynthesis in an ocean biome?

a. aphotic zone

b. abyssal zone
c. benthic realm
d. intertidal zone

Solution:

D

Exercise:

Problem: A key feature of estuaries is

a. low light conditions and high productivity
b. salt water and fresh water
c. frequent algal blooms
d. little or no vegetation

Solution:

B

Free Response

Exercise:

Problem:

Describe the conditions and challenges facing organisms living in the intertidal zone.

Solution:

Organisms living in the intertidal zone must tolerate periodic exposure to air and sunlight and must be able to be periodically dry. They also must be able to endure the pounding waves; for this reason, some

shoreline organisms have hard exoskeletons that provide protection while also reducing the likelihood of drying out.

Glossary

abyssal zone
: the deepest part of the ocean at depths of 4000 m or greater

algal bloom
: a rapid increase of algae in an aquatic system

aphotic zone
: the part of the ocean where photosynthesis cannot occur

benthic realm
: (also, benthic zone) the part of the ocean that extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor

channel
: the bed and banks of a river or stream

coral reef
: an ocean ridge formed by marine invertebrates living in warm shallow waters within the photic zone

cryptofauna
: the invertebrates found within the calcium carbonate substrate of coral reefs

ecosystem services
: the human benefits provided by natural ecosystems

emergent vegetation
: the plants living in bodies of water that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water’s surface

estuary
: a region where fresh water and salt water mix where a river discharges into an ocean or sea

intertidal zone
: the part of the ocean that is closest to land; parts extend above the water at low tide

neritic zone
: the part of the ocean that extends from low tide to the edge of the continental shelf

oceanic zone
: the part of the ocean that begins offshore where the water measures 200 m deep or deeper

pelagic realm
: (also, pelagic zone) the open ocean waters that are not close to the bottom or near the shore

photic zone
: the upper layer of ocean water in which photosynthesis is able to take place

planktivore
: an animal that eats plankton

source water
: the point of origin of a river or stream

wetland
: environment in which the soil is either permanently or periodically saturated with water

Introduction
class="introduction"

Habitat destruction through deforestation, especially of tropical rainforests as seen in this satellite view of Amazon rainforests in Brazil, is a major cause of the current decline in biodiversity. (credit: modification of work by Jesse Allen and Robert Simmon, NASA Earth Observatory)

Biologists estimate that species extinctions are currently 500–1000 times the rate seen previously in Earth's history when there were no unusual geological or climatic events occurring. Biologists call the previous rate the "background" rate of extinction. The current high rates will cause a precipitous decline in the biodiversity (the diversity of species) of the planet in the next century or two. The losses will include many species we know today. Although it is sometimes difficult to predict which species will become extinct, many are listed as endangered (at great risk of extinction). However, the majority of extinctions will be of species that science has not yet even described.

Most of these "invisible" species that will become extinct currently live in tropical rainforests like those of the Amazon basin. These rainforests are the most diverse ecosystems on the planet and are being destroyed rapidly by deforestation, which biologists believe is driving many rare species with

limited distributions extinct. Between 1970 and 2011, almost 20 percent of the Amazon rainforest was lost. Rates are higher in other tropical rainforests. What we are likely to notice on a day-to-day basis as a result of biodiversity loss is that food will be more difficult to produce, clean water will be more difficult to find, and the rate of development of new medicines will become slower, as we depend upon other species for much of these services. This increased loss of biodiversity is almost entirely a result of human activities as we destroy species' habitats, introduce disruptive species into ecosystems, hunt some species to extinction, continue to warm the planet with greenhouse gases, and influence nature in other ways. Slowing the loss of biodiversity is within our abilities if we make dramatic changes in our consumptive behavior and identify and protect the elements of our ecosystems that we depend on for our lives and welfare.

The Conservation of Soil

By the end of this section, you will be able to:

- Describe how soils are formed
- Explain soil composition

Human civilization is dependent upon the planet’s limited soil and water resources. Plants obtain inorganic elements from the soil, which serves as a natural medium for land plants. **Soil** is the outer loose layer that covers the surface of Earth. Soil quality is a major determinant, along with climate, of plant distribution and growth. Soil quality depends not only on the chemical composition of the soil, but also the topography (regional surface features) and the presence of living organisms. In agriculture, the history of the soil, such as the cultivating practices and previous crops, modify the characteristics and fertility of that soil.

Soil develops very slowly over long periods of time, and its formation results from natural and environmental forces acting on mineral, rock, and organic compounds. Soils can be divided into two groups: **organic soils** are those that are formed from sedimentation and primarily composed of organic matter, while those that are formed from the weathering of rocks and are primarily composed of inorganic material are called **mineral soils**. Mineral soils are predominant in terrestrial ecosystems, where soils may be covered by water for part of the year or exposed to the atmosphere.

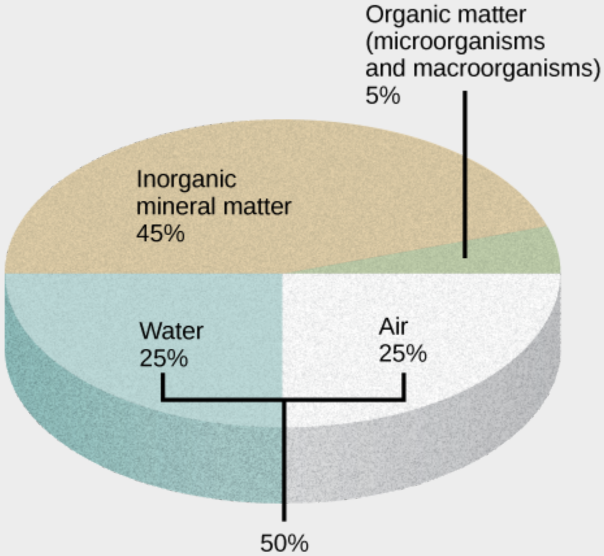
Soil Composition

Soil consists of these major components ([link]):

- inorganic mineral matter, about 40 to 45 percent of the soil volume
- organic matter, about 5 percent of the soil volume
- water and air, about 50 percent of the soil volume

The amount of each of the four major components of soil depends on the amount of vegetation, soil compaction, and water present in the soil. A good healthy soil has sufficient air, water, minerals, and organic material to promote and sustain plant life.

Note:
Art Connection



The four major components of soil are shown: inorganic minerals, organic matter, water, and air.

Soil compaction can result when soil is compressed by heavy machinery or even foot traffic. How might this compaction change the soil composition?

The organic material of soil, called **humus**, is made up of microorganisms (dead and alive), and dead animals and plants in varying stages of decay. Humus improves soil structure and provides plants with water and minerals. The inorganic material of soil consists of rock, slowly broken down into smaller particles that vary in size. Soil particles that are 0.1 to 2 mm in diameter are **sand**. Soil particles between 0.002 and 0.1 mm are called **silt**, and even smaller particles, less than 0.002 mm in diameter, are called **clay**. Some soils have no dominant particle size and contain a mixture of sand, silt, and humus; these soils are called **loams**.

Note:
Link to Learning

Explore this interactive map from the USDA's National Cooperative Soil Survey to access soil data for almost any region in the United States.

Soil Formation

Soil formation is the consequence of a combination of biological, physical, and chemical processes. Soil should ideally contain 50 percent solid material and 50 percent pore space. About one-half of the pore space should contain water, and the other half should contain air. The organic component of soil serves as a cementing agent, returns nutrients to the plant, allows soil to store moisture, makes soil tillable for farming, and provides energy for soil microorganisms. Most soil microorganisms—bacteria, algae, or fungi—are dormant in dry soil, but become active once moisture is available.

Soil distribution is not homogenous because its formation results in the production of layers; together, the vertical section of a soil is called the **soil profile**. Within the soil profile, soil scientists define zones called horizons. A **horizon** is a soil layer with distinct physical and chemical properties that differ from those of other layers. Five factors account for soil formation: parent material, climate, topography, biological factors, and time.

Parent Material

The organic and inorganic material in which soils form is the **parent material**. Mineral soils form directly from the weathering of **bedrock**, the

solid rock that lies beneath the soil, and therefore, they have a similar composition to the original rock. Other soils form in materials that came from elsewhere, such as sand and glacial drift. Materials located in the depth of the soil are relatively unchanged compared with the deposited material. Sediments in rivers may have different characteristics, depending on whether the stream moves quickly or slowly. A fast-moving river could have sediments of rocks and sand, whereas a slow-moving river could have fine-textured material, such as clay.

Climate

Temperature, moisture, and wind cause different patterns of weathering and therefore affect soil characteristics. The presence of moisture and nutrients from weathering will also promote biological activity: a key component of a quality soil.

Topography

Regional surface features (familiarly called “the lay of the land”) can have a major influence on the characteristics and fertility of a soil. Topography affects water runoff, which strips away parent material and affects plant growth. Steeps soils are more prone to erosion and may be thinner than soils that are relatively flat or level.

Biological factors

The presence of living organisms greatly affects soil formation and structure. Animals and microorganisms can produce pores and crevices, and plant roots can penetrate into crevices to produce more fragmentation. Plant secretions promote the development of microorganisms around the root, in an area known as the **rhizosphere**. Additionally, leaves and other material that fall from plants decompose and contribute to soil composition.

Time

Time is an important factor in soil formation because soils develop over long periods. Soil formation is a dynamic process. Materials are deposited over time, decompose, and transform into other materials that can be used by living organisms or deposited onto the surface of the soil.

Section Summary

Plants obtain mineral nutrients from the soil. Soil is the outer loose layer that covers the surface of Earth. Soil quality depends on the chemical composition of the soil, the topography, the presence of living organisms, the climate, and time. Agricultural practice and history may also modify the characteristics and fertility of soil. Soil consists of four major components: 1) inorganic mineral matter, 2) organic matter, 3) water and air, and 4) living matter. The organic material of soil is made of humus, which improves soil structure and provides water and minerals. Soil inorganic material consists of rock slowly broken down into smaller particles that vary in size, such as sand, silt, and loam.

Soil formation results from a combination of biological, physical, and chemical processes. Soil is not homogenous because its formation results in the production of layers called a soil profile. Factors that affect soil formation include: parent material, climate, topography, biological factors, and time. Land conservation is is crucial to sustaining biodiversity of terrestrial plants and animals.

Art Connections

Exercise:

Problem:

[link] Soil compaction can result when soil is compressed by heavy machinery or even foot traffic. How might this compaction change the soil composition?

Solution:

[link] The air content of the soil decreases.

Exercise:

Problem:

[link] Which horizon is considered the topsoil, and which is considered the subsoil?

Solution:

[link] The A horizon is the topsoil, and the B horizon is subsoil.

Review Questions

Exercise:

Problem: Which factors affect soil quality?

a. chemical composition
b. history of the soil
c. presence of living organisms and topography
d. all of the above

Solution:

D

Exercise:

Problem:

Soil particles that are 0.1 to 2 mm in diameter are called _________.

a. sand
b. silt

c. clay
d. loam

Solution:

A

Exercise:

Problem:

A soil consists of layers called _________ that taken together are called a _________.

a. soil profiles : horizon
b. horizons : soil profile
c. horizons : humus
d. humus : soil profile

Solution:

B

Exercise:

Problem:

What is the term used to describe the solid rock that lies beneath the soil?

a. sand
b. bedrock
c. clay
d. loam

Solution:

B

Exercise:

Problem:

Describe the main differences between a mineral soil and an organic soil.

Solution:

A mineral soil forms from the weathering of rocks; it is inorganic material. An organic soil is formed from sedimentation; it mostly consists of humus.

Exercise:

Problem:

Name and briefly explain the factors that affect soil formation.

Solution:

Parent material, climate, topography, biological factors, and time affect soil formation. Parent material is the material in which soils form. Climate describes how temperature, moisture, and wind cause different patterns of weathering, influencing the characteristics of the soil. Topography affects the characteristics and fertility of a soil. Biological factors include the presence of living organisms that greatly affect soil formation. Processes such as freezing and thawing may produce cracks in rocks; plant roots can penetrate these crevices and produce more fragmentation. Time affects soil because soil develops over long periods.

Exercise:

Problem:

Describe how topography influences the characteristics and fertility of a soil.

Solution:

Topography affects water runoff, which strips away parent material and affects plant growth. Steeps soils are more prone to erosion and may be thinner than soils that are on level surfaces.

Glossary

A horizon
: consists of a mixture of organic material with inorganic products of weathering

B horizon
: soil layer that is an accumulation of mostly fine material that has moved downward

bedrock
: solid rock that lies beneath the soil

C horizon
: layer of soil that contains the parent material, and the organic and inorganic material that is broken down to form soil; also known as the soil base

clay
: soil particles that are less than 0.002 mm in diameter

horizon
: soil layer with distinct physical and chemical properties, which differs from other layers depending on how and when it was formed

humus
: organic material of soil; made up of microorganisms, dead animals and plants in varying stages of decay

loam
: soil that has no dominant particle size

mineral soil

type of soil that is formed from the weathering of rocks and inorganic material; composed primarily of sand, silt, and clay

O horizon
: layer of soil with humus at the surface and decomposed vegetation at the base

organic soil
: type of soil that is formed from sedimentation; composed primarily of organic material

parent material
: organic and inorganic material in which soils form

rhizosphere
: area of soil affected by root secretions and microorganisms

sand
: soil particles between 0.1–2 mm in diameter

silt
: soil particles between 0.002 and 0.1 mm in diameter

soil profile
: vertical section of a soil

soil
: outer loose layer that covers the surface of Earth

Importance of Biodiversity
By the end of this section, you will be able to:

- Describe biodiversity as the equilibrium of naturally fluctuating rates of extinction and speciation
- Identify benefits of biodiversity to humans

This tropical lowland rainforest in Madagascar is an example of a high biodiversity habitat. This particular location is protected within a national forest, yet only 10 percent of the original coastal lowland forest remains, and research suggests half the original biodiversity has been lost. (credit: Frank Vassen)

Biodiversity is a broad term for biological variety, and it can be measured at a number of organizational levels. Traditionally, ecologists have measured **biodiversity** by taking into account both the number of species and the number of individuals in each of those species. However, biologists are using measures of biodiversity at several levels of biological organization (including genes, populations, and ecosystems) to help focus efforts to preserve the biologically and technologically important elements of biodiversity.

When biodiversity loss through extinction is thought of as the loss of the passenger pigeon, the dodo, or, even, the woolly mammoth there seems to be no reason to care about it because these events happened long ago. How is the loss practically important for the welfare of the human species? Would these species have made our lives any better? From the perspective of evolution and ecology, the loss of a particular individual species, with some exceptions, may seem unimportant, but the current accelerated extinction rate means the loss of tens of thousands of species within our lifetimes. Much of this loss is occurring in tropical rainforests like the one pictured in [link], which are especially high-diversity ecosystems that are being cleared for timber and agriculture. This is likely to have dramatic effects on human welfare through the collapse of ecosystems and in added costs to maintain food production, clean air and water, and improve human health.

Biologists recognize that human populations are embedded in ecosystems and are dependent on them, just as is every other species on the planet. Agriculture began after early hunter-gatherer societies first settled in one place and heavily modified their immediate environment: the ecosystem in which they existed. This cultural transition has made it difficult for humans to recognize their dependence on living things other than crops and domesticated animals on the planet. Today our technology smoothes out the extremes of existence and allows many of us to live longer, more comfortable lives, but ultimately the human species cannot exist without its surrounding ecosystems. Our ecosystems provide our food. This includes living plants that grow in soil ecosystems and the animals that eat these plants (or other animals) as well as photosynthetic organisms in the oceans and the other organisms that eat them. Our ecosystems have provided and will provide many of the medications that maintain our health, which are

commonly made from compounds found in living organisms. Ecosystems provide our clean water, which is held in lake and river ecosystems or passes through terrestrial ecosystems on its way into groundwater.

Types of Biodiversity

A common meaning of biodiversity is simply the number of species in a location or on Earth; for example, the American Ornithologists' Union lists 2078 species of birds in North and Central America. This is one measure of the bird biodiversity on the continent. More sophisticated measures of diversity take into account the relative abundances of species. For example, a forest with 10 equally common species of trees is more diverse than a forest that has 10 species of trees wherein just one of those species makes up 95 percent of the trees rather than them being equally distributed. Biologists have also identified alternate measures of biodiversity, some of which are important in planning how to preserve biodiversity.

Genetic and Chemical Biodiversity

Genetic diversity is one alternate concept of biodiversity. **Genetic diversity** (or variation) is the raw material for adaptation in a species. A species' future potential for adaptation depends on the genetic diversity held in the genomes of the individuals in populations that make up the species. The same is true for higher taxonomic categories. A genus with very different types of species will have more genetic diversity than a genus with species that look alike and have similar ecologies. The genus with the greatest potential for subsequent evolution is the most genetically diverse one.

Most genes code for proteins, which in turn carry out the metabolic processes that keep organisms alive and reproducing. Genetic diversity can also be conceived of as **chemical diversity** in that species with different genetic makeups produce different assortments of chemicals in their cells (proteins as well as the products and byproducts of metabolism). This chemical diversity is important for humans because of the potential uses for these chemicals, such as medications. For example, the drug eptifibatide is derived from rattlesnake venom and is used to prevent heart attacks in individuals with certain heart conditions.

At present, it is far cheaper to discover compounds made by an organism than to imagine them and then synthesize them in a laboratory. Chemical diversity is one way to measure diversity that is important to human health and welfare. Through selective breeding, humans have domesticated animals, plants, and fungi, but even this diversity is suffering losses because of market forces and increasing globalism in human agriculture and migration. For example, international seed companies produce only a very few varieties of a given crop and provide incentives around the world for farmers to buy these few varieties while abandoning their traditional varieties, which are far more diverse. The human population depends on crop diversity directly as a stable food source and its decline is troubling to biologists and agricultural scientists.

Ecosystems Diversity

It is also useful to define **ecosystem diversity**: the number of different ecosystems on Earth or in a geographical area. Whole ecosystems can disappear even if some of the species might survive by adapting to other ecosystems. The loss of an ecosystem means the loss of the interactions between species, the loss of unique features of coadaptation, and the loss of biological productivity that an ecosystem is able to create. An example of a largely extinct ecosystem in North America is the prairie ecosystem ([link]). Prairies once spanned central North America from the boreal forest in northern Canada down into Mexico. They are now all but gone, replaced by crop fields, pasture lands, and suburban sprawl. Many of the species survive, but the hugely productive ecosystem that was responsible for creating our most productive agricultural soils is now gone. As a consequence, their soils are now being depleted unless they are maintained artificially at greater expense. The decline in soil productivity occurs because the interactions in the original ecosystem have been lost; this was a far more important loss than the relatively few species that were driven extinct when the prairie ecosystem was destroyed.

The variety of ecosystems on Earth—from coral reef to prairie—enables a great diversity of species to exist. (credit "coral reef": modification of work by Jim Maragos, USFWS; credit: "prairie": modification of work by Jim Minnerath, USFWS)

Current Species Diversity

Despite considerable effort, knowledge of the species that inhabit the planet is limited. A recent estimate suggests that the eukaryote species for which science has names, about 1.5 million species, account for less than 20 percent of the total number of eukaryote species present on the planet (8.7 million species, by one estimate). Estimates of numbers of prokaryotic species are largely guesses, but biologists agree that science has only just begun to catalog their diversity. Even with what is known, there is no centralized repository of names or samples of the described species; therefore, there is no way to be sure that the 1.5 million descriptions is an accurate number. It is a best guess based on the opinions of experts on different taxonomic groups. Given that Earth is losing species at an accelerating pace, science knows little about what is being lost. [link] presents recent estimates of biodiversity in different groups.

Estimated Numbers of Described and Predicted species

	Source: Mora et al 2011		**Source: Chapman 2009**		**Source: Groombridge and Jenkins 2002**	
	Described	Predicted	Described	Predicted	Described	Predicted
Animals	1,124,516	9,920,000	1,424,153	6,836,330	1,225,500	10,820,000
Photosynthetic protists	17,892	34,900	25,044	200,500	—	—
Fungi	44,368	616,320	98,998	1,500,000	72,000	1,500,000
Plants	224,244	314,600	310,129	390,800	270,000	320,000
Non-photosynthetic protists	16,236	72,800	28,871	1,000,000	80,000	600,000
Prokaryotes	—	—	10,307	1,000,000	10,175	—

Estimated Numbers of Described and Predicted species						
	Source: Mora et al 2011		**Source: Chapman 2009**		**Source: Groombridge and Jenkins 2002**	
Total	1,438,769	10,960,000	1,897,502	10,897,630	1,657,675	13,240,000

This table shows the estimated number of species by taxonomic group—including both described (named and studied) and predicted (yet to be named) species.

There are various initiatives to catalog described species in accessible and more organized ways, and the internet is facilitating that effort. Nevertheless, at the current rate of species description, which according to the State of Observed Species[footnote] reports is 17,000–20,000 new species a year, it would take close to 500 years to describe all of the species currently in existence. The task, however, is becoming increasingly impossible over time as **extinction** removes species from Earth faster than they can be described.
International Institute for Species Exploration (IISE), *2011 State of Observed Species (SOS)*. Tempe, AZ: IISE, 2011. Accessed May, 20, 2012. http://species.asu.edu/SOS.

Naming and counting species may seem an unimportant pursuit given the other needs of humanity, but it is not simply an accounting. Describing species is a complex process by which biologists determine an organism's unique characteristics and whether or not that organism belongs to any other described species. It allows biologists to find and recognize the species after the initial discovery to follow up on questions about its biology. That subsequent research will produce the discoveries that make the species valuable to humans and to our ecosystems. Without a name and description, a species cannot be studied in depth and in a coordinated way by multiple scientists.

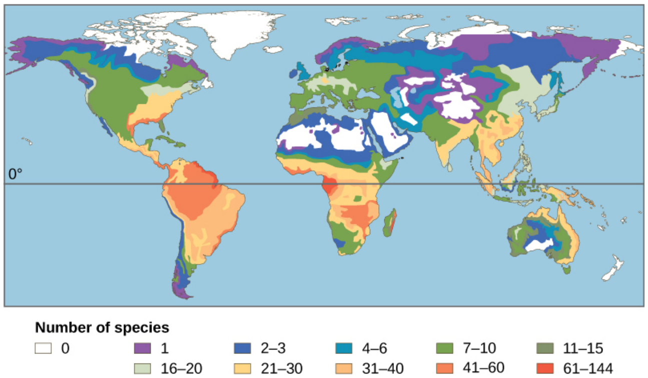
Patterns of Biodiversity

Biodiversity is not evenly distributed on the planet. Lake Victoria contained almost 500 species of cichlids (only one family of fishes present in the lake) before the introduction of an exotic species in the 1980s and 1990s caused a mass extinction. All of these species were found only in Lake Victoria, which is to say they were endemic. **Endemic species** are found in only one location. For example, the blue jay is endemic to North America, while the Barton Springs salamander is endemic to the mouth of one spring in Austin, Texas. Endemics with highly restricted distributions, like the Barton Springs salamander, are particularly vulnerable to extinction. Higher taxonomic levels, such as genera and families, can also be endemic.

Lake Huron contains about 79 species of fish, all of which are found in many other lakes in North America. What accounts for the difference in diversity between Lake Victoria and Lake Huron? Lake Victoria is a tropical lake, while Lake Huron is a temperate lake. Lake Huron in its present form is only about 7,000 years old, while Lake Victoria in its present form is about 15,000 years old. These two factors, latitude and age, are two of several hypotheses biogeographers have suggested to explain biodiversity patterns on Earth.

Note:
Career in Action
Biogeography
Biogeography is the study of the distribution of the world's species both in the past and in the present. The work of biogeographers is critical to understanding our physical environment, how the environment affects species, and how changes in environment impact the distribution of a species.
There are three main fields of study under the heading of biogeography: ecological biogeography, historical biogeography (called paleobiogeography), and conservation biogeography. Ecological biogeography studies the current factors affecting the distribution of plants and animals. Historical biogeography, as the name implies, studies the past distribution of species. Conservation biogeography, on the other hand, is focused on the protection and restoration of species based upon the known historical and current ecological information. Each of these fields considers both zoogeography and phytogeography—the past and present distribution of animals and plants.

One of the oldest observed patterns in ecology is that biodiversity in almost every taxonomic group of organism increases as latitude declines. In other words, biodiversity increases closer to the equator ([link]).



This map illustrates the number of amphibian species across the globe and shows the trend toward higher biodiversity at lower latitudes. A similar pattern is observed for most taxonomic groups.

It is not yet clear why biodiversity increases closer to the equator, but hypotheses include the greater age of the ecosystems in the tropics versus temperate regions, which were largely devoid of life or drastically impoverished during the last ice age. The greater age provides more time for speciation. Another possible explanation is the greater energy the tropics receive from the sun versus the lesser energy input in temperate and polar regions. But scientists have not been able to explain how greater energy input could translate into more species. The complexity of tropical ecosystems may promote speciation by increasing the **habitat heterogeneity**, or number of ecological niches, in the tropics relative to higher latitudes. The greater heterogeneity provides more opportunities for coevolution, specialization, and perhaps greater selection pressures leading to population differentiation. However, this hypothesis suffers from some circularity—ecosystems with more species encourage speciation, but how did they get more species to begin with? The tropics have been perceived as being more stable than temperate regions, which have a pronounced climate and day-length seasonality. The tropics have their own forms of seasonality, such as rainfall, but they are generally assumed to be more stable environments and this stability might promote speciation.

Regardless of the mechanisms, it is certainly true that biodiversity is greatest in the tropics. The number of endemic species is higher in the tropics. The tropics also contain more biodiversity hotspots. At the same time, our knowledge of the species living in the tropics is lowest and because of recent, heavy human activity the potential for biodiversity loss is greatest.

Importance of Biodiversity

Loss of biodiversity eventually threatens other species we do not impact directly because of their interconnectedness; as species disappear from an ecosystem other species are threatened by the changes in available resources. Biodiversity is important to the survival and welfare of human populations because it has impacts on our health and our ability to feed ourselves through agriculture and harvesting populations of wild animals.

Human Health

Many medications are derived from natural chemicals made by a diverse group of organisms. For example, many plants produce **secondary plant compounds**, which are toxins used to protect the plant from insects and other animals that eat them. Some of these secondary plant compounds also work as human medicines. Contemporary societies that live close to the land often have a broad knowledge of the medicinal uses of plants growing in their area. For centuries in Europe, older knowledge about the medical uses of plants was compiled in herbals—books that identified the plants and their uses. Humans are not the only animals to use plants for medicinal reasons. The other great apes, orangutans, chimpanzees, bonobos, and gorillas have all been observed self-medicating with plants.

Modern pharmaceutical science also recognizes the importance of these plant compounds. Examples of significant medicines derived from plant compounds include aspirin, codeine, digoxin, atropine, and vincristine ([link]). Many medications were once derived from plant extracts but are now synthesized. It is estimated that, at one time, 25 percent of modern drugs contained at least one plant extract. That number has probably decreased to about 10 percent as natural plant ingredients are replaced by synthetic versions of the plant compounds. Antibiotics, which are responsible for extraordinary improvements in health and lifespans in developed countries, are compounds largely derived from fungi and bacteria.

Catharanthus roseus, the Madagascar periwinkle, has various medicinal properties. Among other uses, it is a source of vincristine, a drug used in the treatment of lymphomas. (credit: Forest and Kim Starr)

In recent years, animal venoms and poisons have excited intense research for their medicinal potential. By 2007, the FDA had approved five drugs based on animal toxins to treat diseases such as hypertension, chronic pain, and diabetes. Another five drugs are undergoing clinical trials and at least six drugs are being used in other countries. Other toxins under investigation come from mammals, snakes, lizards, various amphibians, fish, snails, octopuses, and scorpions.

Aside from representing billions of dollars in profits, these medications improve people’s lives. Pharmaceutical companies are actively looking for new natural compounds that can function as medicines. It is estimated that one third of pharmaceutical research and development is spent on natural compounds and that about 35 percent of new drugs brought to market between 1981 and 2002 were from natural compounds.

Finally, it has been argued that humans benefit psychologically from living in a biodiverse world. The chief proponent of this idea is entomologist E. O. Wilson. He argues that human evolutionary history has adapted us to living in a natural environment and that built environments generate stresses that affect human health and well-being. There is considerable research into the psychologically regenerative benefits of natural landscapes that suggest the hypothesis may hold some truth.

Agricultural

Since the beginning of human agriculture more than 10,000 years ago, human groups have been breeding and selecting crop varieties. This crop diversity matched the cultural diversity of highly subdivided populations of humans. For example, potatoes were domesticated beginning around 7,000 years ago in the central Andes of Peru and Bolivia. The people in this region traditionally lived in relatively isolated settlements separated by mountains. The potatoes grown in that region belong to seven species and the number of varieties likely is in the thousands. Each variety has been bred to thrive at particular elevations and soil and climate conditions. The diversity is driven by the diverse demands of the dramatic elevation changes, the limited movement of people, and the demands created by crop rotation for different varieties that will do well in different fields.

Potatoes are only one example of agricultural diversity. Every plant, animal, and fungus that has been cultivated by humans has been bred from original wild ancestor species into diverse varieties arising from the demands for food value, adaptation to growing conditions, and resistance to pests. The potato demonstrates a well-known example of the risks of low crop diversity: during the tragic Irish potato famine (1845–1852 AD), the single potato variety grown in Ireland became susceptible to a potato blight—wiping out the crop. The loss of the crop led to famine, death, and mass emigration. Resistance to disease is a chief benefit to maintaining crop biodiversity and lack of diversity in contemporary crop species carries similar risks. Seed companies, which are the source of most crop varieties in developed countries, must continually breed new varieties to keep up with evolving pest organisms. These same seed companies, however, have participated in the decline of the number of varieties available as they focus on selling fewer varieties in more areas of the world replacing traditional local varieties.

The ability to create new crop varieties relies on the diversity of varieties available and the availability of wild forms related to the crop plant. These wild forms are often the source of new gene variants that can be bred with existing varieties to create varieties with new attributes. Loss of wild species related to a crop will mean the loss of potential in crop improvement. Maintaining the genetic diversity of wild species related to domesticated species ensures our continued supply of food.

Since the 1920s, government agriculture departments have maintained seed banks of crop varieties as a way to maintain crop diversity. This system has flaws because over time seed varieties are lost through accidents and there is no way to replace them. In 2008, the Svalbard Global seed Vault, located on Spitsbergen island, Norway, ([link]) began storing seeds from around the world as a backup system to the regional seed banks. If a regional seed bank stores varieties in Svalbard, losses can be replaced from Svalbard should something happen to the regional seeds. The Svalbard seed vault is deep into the rock of the arctic island. Conditions within the vault are maintained at ideal temperature and humidity for seed survival, but the deep underground location of the vault in the arctic means that failure of the vault's systems will not compromise the climatic conditions inside the vault.

Note:
Art Connection

The Svalbard Global Seed Vault is a storage facility for seeds of Earth's diverse crops. (credit: Mari Tefre, Svalbard Global Seed Vault)

The Svalbard seed vault is located on Spitsbergen island in Norway, which has an arctic climate. Why might an arctic climate be good for seed storage?

Although crops are largely under our control, our ability to grow them is dependent on the biodiversity of the ecosystems in which they are grown. That biodiversity creates the conditions under which crops are able to grow through what are known as ecosystem services—valuable conditions or processes that are carried out by an ecosystem. Crops are not grown, for the most part, in built environments. They are grown in soil. Although some agricultural soils are rendered sterile using controversial pesticide treatments, most contain a huge diversity of organisms that maintain nutrient cycles—breaking down organic matter into nutrient compounds that crops need for growth. These organisms also maintain soil texture that affects water and oxygen dynamics in the soil that are necessary for plant growth. Replacing the work of these organisms in forming arable soil is not practically possible. These kinds of processes are called ecosystem services. They occur within ecosystems, such as soil ecosystems, as a result of the diverse metabolic activities of the organisms living there, but they provide benefits to human food production, drinking water availability, and breathable air.

Other key ecosystem services related to food production are plant pollination and crop pest control. It is estimated that honeybee pollination within the United States brings in $1.6 billion per year; other pollinators contribute up to $6.7 billion. Over 150 crops in the United States require pollination to produce. Many honeybee populations are managed by beekeepers who rent out their hives' services to farmers. Honeybee populations in North America have been suffering large losses caused by a syndrome known as colony collapse disorder, a new phenomenon with an unclear cause. Other pollinators include a diverse array of other bee species and various insects and birds. Loss of these species would make growing crops requiring pollination impossible, increasing dependence on other crops.

Finally, humans compete for their food with crop pests, most of which are insects. Pesticides control these competitors, but these are costly and lose their effectiveness over time as pest populations adapt. They also lead to collateral damage by killing non-pest species as well as beneficial insects like honeybees, and risking the health of agricultural workers and consumers. Moreover, these pesticides may migrate from the fields where they are applied and do damage to other ecosystems like streams, lakes, and even the ocean. Ecologists believe that the bulk of the work in removing pests is actually done by predators and parasites of those pests, but the impact has not been well studied. A review found that in 74 percent of studies that looked for an effect of landscape complexity (forests and fallow fields near to crop fields) on natural enemies of pests, the greater the complexity, the greater the effect of pest-suppressing organisms. Another experimental study found that introducing multiple enemies of pea aphids (an important alfalfa pest) increased the yield of alfalfa significantly. This study shows that a diversity of pests is more effective at control than one single pest. Loss of diversity in pest enemies will inevitably make it more difficult and costly to grow food. The world's growing human population faces significant challenges in the increasing costs and other difficulties associated with producing food.

Wild Food Sources

In addition to growing crops and raising food animals, humans obtain food resources from wild populations, primarily wild fish populations. For about one billion people, aquatic resources provide the main source of animal protein. But since 1990, production from global fisheries has declined. Despite considerable effort, few fisheries on Earth are managed sustainability.

Fishery extinctions rarely lead to complete extinction of the harvested species, but rather to a radical restructuring of the marine ecosystem in which a dominant species is so over-harvested that it becomes a minor player, ecologically. In addition to humans losing the food source, these alterations affect many other species in ways that are difficult or impossible to predict. The collapse of fisheries has dramatic and long-lasting effects on local human populations that work in the fishery. In addition, the loss of an inexpensive protein source to populations that cannot afford to replace it will increase the cost of living and limit societies in other ways. In general, the fish

taken from fisheries have shifted to smaller species and the larger species are overfished. The ultimate outcome could clearly be the loss of aquatic systems as food sources.

Note:
Concept in Action

Visit this website to view a brief video discussing a study of declining fisheries.

Section Summary

Biodiversity exists at multiple levels of organization, and is measured in different ways depending on the goals of those taking the measurements. These include numbers of species, genetic diversity, chemical diversity, and ecosystem diversity. The number of described species is estimated to be 1.5 million with about 17,000 new species being described each year. Estimates for the total number of eukaryotic species on Earth vary but are on the order of 10 million. Biodiversity is negatively correlated with latitude for most taxa, meaning that biodiversity is higher in the tropics. The mechanism for this pattern is not known with certainty, but several plausible hypotheses have been advanced.

Humans use many compounds that were first discovered or derived from living organisms as medicines: secondary plant compounds, animal toxins, and antibiotics produced by bacteria and fungi. More medicines are expected to be discovered in nature. Loss of biodiversity will impact the number of pharmaceuticals available to humans. Biodiversity may provide important psychological benefits to humans.

Crop diversity is a requirement for food security, and it is being lost. The loss of wild relatives to crops also threatens breeders' abilities to create new varieties. Ecosystems provide ecosystem services that support human agriculture: pollination, nutrient cycling, pest control, and soil development and maintenance. Loss of biodiversity threatens these ecosystem services and risks making food production more expensive or impossible. Wild food sources are mainly aquatic, but few are being managed for sustainability. Fisheries' ability to provide protein to human populations is threatened when extinction occurs.

Art Connections

Exercise:

Problem:

[link] The Svalbard seed vault is located on Spitsbergen island in Norway, which has an arctic climate. Why might an arctic climate be good for seed storage?

Solution:

[link] The ground is permanently frozen so the seeds will keep, even if the electricity fails.

Multiple Choice

Exercise:

Problem: The number of currently described species on the planet is about ________.

a. 17,000
b. 150,000
c. 1.5 million
d. 10 million

Solution:

C

Exercise:

Problem:A secondary plant compound might be used for which of the following?

a. a new crop variety
b. a new drug
c. a soil nutrient
d. a crop pest

Solution:

B

Exercise:

Problem: Pollination is an example of ________.

a. a possible source of new drugs
b. chemical diversity
c. an ecosystem service
d. crop pest control

Solution:

C

Free Response

Exercise:

Problem: Explain how biodiversity loss can impact crop diversity.

Solution:

Crop plants are derived from wild plants, and genes from wild relatives are frequently brought into crop varieties by plant breeders to add valued characteristics to the crops. If the wild species are lost, then this genetic variation would no longer be available.

Exercise:

Problem: Describe two types of compounds from living things that are used as medications.

Solution:

Secondary plant compounds are toxins produced by plants to kill predators trying to eat them; some of these compounds can be used as drugs. Animal toxins, such as snake venom, can be used as medicine. (Alternate answer: antibiotics are compounds produced by bacteria and fungi which can be used to kill bacteria.)

Glossary

biodiversity
: the variety of a biological system, typically conceived as the number of species, but also applying to genes, biochemistry, and ecosystems

chemical diversity
: the variety of metabolic compounds in an ecosystem

ecosystem diversity
: the variety of ecosystems

endemic species
: a species native to one place

extinction
: the disappearance of a species from Earth; local extinction is the disappearance of a species from a region

genetic diversity
: the variety of genes and alleles in a species or other taxonomic group or ecosystem; the term can refer to allelic diversity or genome-wide diversity

habitat heterogeneity
: the number of ecological niches

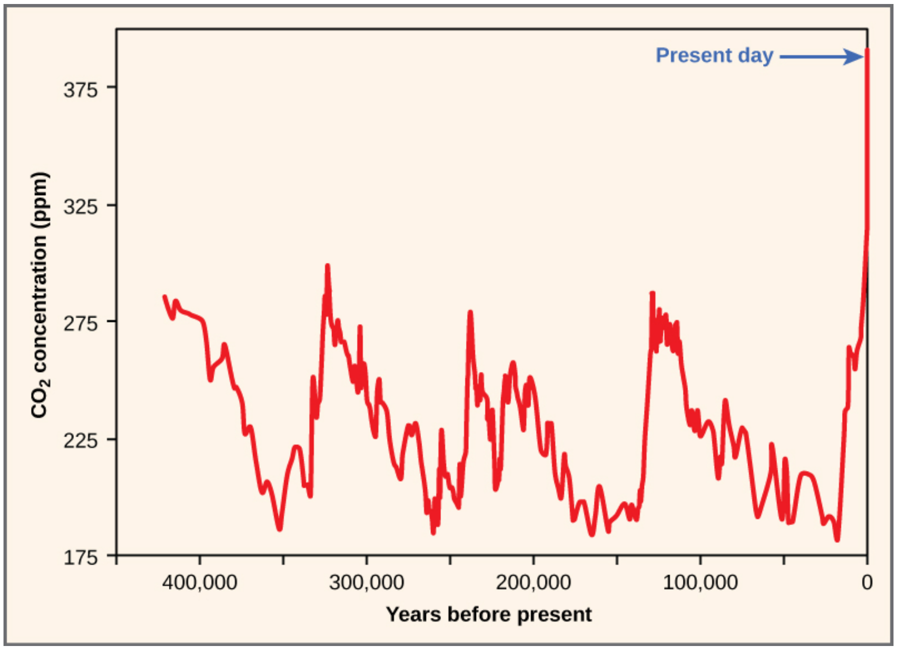
secondary plant compound
: a compound produced as a byproduct of plant metabolic processes that is typically toxic, but is sequestered by the plant to defend against herbivores

Threats to Biodiversity

By the end of this section, you will be able to:

- Identify significant threats to biodiversity
- Explain the effects of habitat loss, exotic species, and hunting on biodiversity
- Identify the early and predicted effects of climate change on biodiversity

The core threat to biodiversity on the planet, and therefore a threat to human welfare, is the combination of human population growth and the resources used by that population. The human population requires resources to survive and grow, and those resources are being removed unsustainably from the environment. The three greatest proximate threats to biodiversity are habitat loss, overharvesting, and introduction of exotic species. The first two of these are a direct result of human population growth and resource use. The third results from increased mobility and trade. A fourth major cause of extinction, anthropogenic (human-caused) climate change, has not yet had a large impact, but it is predicted to become significant during this century. Global climate change is also a consequence of human population needs for energy and the use of fossil fuels to meet those needs ([link]). Environmental issues, such as toxic pollution, have specific targeted effects on species, but are not generally seen as threats at the magnitude of the others.



Atmospheric carbon dioxide levels fluctuate in a cyclical manner. However, the burning of fossil fuels in recent history has caused a dramatic increase in the levels of carbon dioxide in the Earth's atmosphere, which have now reached levels never before seen on Earth. Scientists predict that the addition of this "greenhouse gas" to the atmosphere is resulting in climate change that will significantly impact biodiversity in the coming century.

Habitat Loss

Humans rely on technology to modify their environment and replace certain functions that were once performed by the natural ecosystem. Other species cannot do this. Elimination of their habitat—whether it is a forest, coral reef, grassland, or flowing river—will kill the individuals in the species. Remove the entire habitat within the range of a species and, unless they are one of the few species that do well in human-built environments, the species will become extinct. Human destruction of habitats (habitats

generally refer to the part of the ecosystem required by a particular species) accelerated in the latter half of the twentieth century. Consider the exceptional biodiversity of Sumatra: it is home to one species of orangutan, a species of critically endangered elephant, and the Sumatran tiger, but half of Sumatra’s forest is now gone. The neighboring island of Borneo, home to the other species of orangutan, has lost a similar area of forest. Forest loss continues in protected areas of Borneo. The orangutan in Borneo is listed as endangered by the International Union for Conservation of Nature (IUCN), but it is simply the most visible of thousands of species that will not survive the disappearance of the forests of Borneo. The forests are removed for timber and to plant palm oil plantations ([link]). Palm oil is used in many products including food products, cosmetics, and biodiesel in Europe. A 5-year estimate of global forest cover loss for the years from 2000 to 2005 was 3.1 percent. Much loss (2.4 percent) occurred in the humid tropics where forest loss is primarily from timber extraction. These losses certainly also represent the extinction of species unique to those areas.

An oil palm plantation in Sabah province Borneo, Malaysia, replaces native forest habitat that a variety of species depended on to live. (credit: Lian Pin Koh)

Note:
Biology in Action
Preventing Habitat Destruction with Wise Wood Choices
Most consumers do not imagine that the home improvement products they buy might be contributing to habitat loss and species extinctions. Yet the market for illegally harvested tropical timber is huge, and the wood products often find themselves in building supply stores in the United States. One estimate is that 10 percent of the imported timber stream in the United States, which is the world's largest consumer of wood products, is potentially illegally logged. In 2006, this amounted to $3.6 billion in wood products. Most of the illegal products are imported from countries that act as intermediaries and are not the originators of the wood.
How is it possible to determine if a wood product, such as flooring, was harvested sustainably or even legally? The Forest Stewardship Council (FSC) certifies sustainably harvested forest products; therefore, looking for their certification on flooring and other hardwood products is one way to ensure that the wood has not been taken illegally from a tropical forest. Certification applies to specific products, not to a producer; some producers' products may not have certification while other products are certified. There are certifications other than the FSC, but these are run by timber companies creating a conflict of interest. Another approach is to buy domestic wood species. While it would be great if there was a list of legal versus illegal woods, it is not that simple. Logging and forest management laws vary from country to country; what is illegal in one country may be legal in another. Where and how a product is harvested and whether the forest from which it comes is being sustainably maintained all factor into whether a wood product will be certified by the FSC. It is always a good idea to ask questions about where a wood product came from and how the supplier knows that it was harvested legally.

Habitat destruction can affect ecosystems other than forests. Rivers and streams are important ecosystems and are frequently the target of habitat modification through building and from damming or water removal. Damming of rivers affects flows and access to all parts of a river. Altering a flow regime can reduce or eliminate populations that are adapted to

seasonal changes in flow. For example, an estimated 91 percent of river lengths in the United States have been modified with damming or bank modifications. Many fish species in the United States, especially rare species or species with restricted distributions, have seen declines caused by river damming and habitat loss. Research has confirmed that species of amphibians that must carry out parts of their life cycles in both aquatic and terrestrial habitats are at greater risk of population declines and extinction because of the increased likelihood that one of their habitats or access between them will be lost. This is of particular concern because amphibians have been declining in numbers and going extinct more rapidly than many other groups for a variety of possible reasons.

Overharvesting

Overharvesting is a serious threat to many species, but particularly to aquatic species. There are many examples of regulated fisheries (including hunting of marine mammals and harvesting of crustaceans and other species) monitored by fisheries scientists that have nevertheless collapsed. The western Atlantic cod fishery is the most spectacular recent collapse. While it was a hugely productive fishery for 400 years, the introduction of modern factory trawlers in the 1980s and the pressure on the fishery led to it becoming unsustainable. The causes of fishery collapse are both economic and political in nature. Most fisheries are managed as a common resource, available to anyone willing to fish, even when the fishing territory lies within a country’s territorial waters. Common resources are subject to an economic pressure known as the **tragedy of the commons**, in which fishers have little motivation to exercise restraint in harvesting a fishery when they do not own the fishery. The general outcome of harvests of resources held in common is their overexploitation. While large fisheries are regulated to attempt to avoid this pressure, it still exists in the background. This overexploitation is exacerbated when access to the fishery is open and unregulated and when technology gives fishers the ability to overfish. In a few fisheries, the biological growth of the resource is less than the potential growth of the profits made from fishing if that time and money were invested elsewhere. In these cases—whales are an example—economic forces will drive toward fishing the population to extinction.

Note:
Concept in Action

Explore a U.S. Fish & Wildlife Service interactive map of critical habitat for endangered and threatened species in the United States. To begin, select “Visit the online mapper.”

For the most part, fishery extinction is not equivalent to biological extinction—the last fish of a species is rarely fished out of the ocean. But there are some instances in which true extinction is a possibility. Whales have slow-growing populations and are at risk of complete extinction through hunting. Also, there are some species of sharks with restricted distributions that are at risk of extinction. The groupers are another population of generally slow-growing fishes that, in the Caribbean, includes a number of species that are at risk of extinction from overfishing.

Coral reefs are extremely diverse marine ecosystems that face peril from several processes. Reefs are home to 1/3 of the world’s marine fish species—about 4000 species—despite making up only one percent of marine habitat. Most home marine aquaria house coral reef species that are wild-caught organisms—not cultured organisms. Although no marine species is known to have been driven extinct by the pet trade, there are studies showing that populations of some species have declined in response to harvesting, indicating that the harvest is not sustainable at those levels. There are also concerns about the effect of the pet trade on some terrestrial species such as turtles, amphibians, birds, plants, and even the orangutans.

Note:

Concept in Action

View a brief video discussing the role of marine ecosystems in supporting human welfare and the decline of ocean ecosystems.

Bush meat is the generic term used for wild animals killed for food. Hunting is practiced throughout the world, but hunting practices, particularly in equatorial Africa and parts of Asia, are believed to threaten several species with extinction. Traditionally, bush meat in Africa was hunted to feed families directly; however, recent commercialization of the practice now has bush meat available in grocery stores, which has increased harvest rates to the level of unsustainability. Additionally, human population growth has increased the need for protein foods that are not being met from agriculture. Species threatened by the bush meat trade are mostly mammals including many monkeys and the great apes living in the Congo basin.

Exotic Species

Exotic species are species that have been intentionally or unintentionally introduced by humans into an ecosystem in which they did not evolve. Human transportation of people and goods, including the intentional transport of organisms for trade, has dramatically increased the introduction of species into new ecosystems. These new introductions are sometimes at distances that are well beyond the capacity of the species to ever travel itself and outside the range of the species' natural predators.

Most exotic species introductions probably fail because of the low number of individuals introduced or poor adaptation to the ecosystem they enter.

Some species, however, have characteristics that can make them especially successful in a new ecosystem. These exotic species often undergo dramatic population increases in their new habitat and reset the ecological conditions in the new environment, threatening the species that exist there. When this happens, the exotic species also becomes an invasive species. Invasive species can threaten other species through competition for resources, predation, or disease.

Note:
Concept in Action

Explore this interactive global database of exotic or invasive species.

Lakes and islands are particularly vulnerable to extinction threats from introduced species. In Lake Victoria, the intentional introduction of the Nile perch was largely responsible for the extinction of about 200 species of cichlids. The accidental introduction of the brown tree snake via aircraft ([link]) from the Solomon Islands to Guam in 1950 has led to the extinction of three species of birds and three to five species of reptiles endemic to the island. Several other species are still threatened. The brown tree snake is adept at exploiting human transportation as a means to migrate; one was even found on an aircraft arriving in Corpus Christi, Texas. Constant vigilance on the part of airport, military, and commercial aircraft personnel is required to prevent the snake from moving from Guam to other islands in the Pacific, especially Hawaii. Islands do not make up a large area of land on the globe, but they do contain a disproportionate number of endemic species because of their isolation from mainland ancestors.

The brown tree snake, *Boiga irregularis*, is an exotic species that has caused numerous extinctions on the island of Guam since its accidental introduction in 1950. (credit: NPS)

Many introductions of aquatic species, both marine and freshwater, have occurred when ships have dumped ballast water taken on at a port of origin into waters at a destination port. Water from the port of origin is pumped into tanks on a ship empty of cargo to increase stability. The water is drawn from the ocean or estuary of the port and typically contains living organisms such as plant parts, microorganisms, eggs, larvae, or aquatic animals. The water is then pumped out before the ship takes on cargo at the destination port, which may be on a different continent. The zebra mussel was introduced to the Great Lakes from Europe prior to 1988 in ship ballast. The zebra mussels in the Great Lakes have cost the industry millions of dollars in clean up costs to maintain water intakes and other facilities. The mussels have also altered the ecology of the lakes dramatically. They threaten native mollusk populations, but have also benefited some species, such as smallmouth bass. The mussels are filter feeders and have dramatically improved water clarity, which in turn has allowed aquatic plants to grow along shorelines, providing shelter for young

fish where it did not exist before. The European green crab, *Carcinus maenas*, was introduced to San Francisco Bay in the late 1990s, likely in ship ballast water, and has spread north along the coast to Washington. The crabs have been found to dramatically reduce the abundance of native clams and crabs with resulting increases in the prey of native crabs.

Invading exotic species can also be disease organisms. It now appears that the global decline in amphibian species recognized in the 1990s is, in some part, caused by the fungus *Batrachochytrium dendrobatidis*, which causes the disease **chytridiomycosis** ([link]). There is evidence that the fungus is native to Africa and may have been spread throughout the world by transport of a commonly used laboratory and pet species: the African clawed frog, *Xenopus laevis*. It may well be that biologists themselves are responsible for spreading this disease worldwide. The North American bullfrog, *Rana catesbeiana*, which has also been widely introduced as a food animal but which easily escapes captivity, survives most infections of *B. dendrobatidis* and can act as a reservoir for the disease.

This Limosa harlequin frog (*Atelopus limosus*), an endangered species from Panama, died from a

fungal disease called chytridiomycosis. The red lesions are symptomatic of the disease. (credit: Brian Gratwicke)

Early evidence suggests that another fungal pathogen, *Geomyces destructans*, introduced from Europe is responsible for **white-nose syndrome**, which infects cave-hibernating bats in eastern North America and has spread from a point of origin in western New York State ([link]). The disease has decimated bat populations and threatens extinction of species already listed as endangered: the Indiana bat, *Myotis sodalis*, and potentially the Virginia big-eared bat, *Corynorhinus townsendii virginianus*. How the fungus was introduced is unknown, but one logical presumption would be that recreational cavers unintentionally brought the fungus on clothes or equipment from Europe.

This little brown bat in Greeley Mine, Vermont, March 26, 2009, was

found to have white-nose syndrome. (credit: modification of work by Marvin Moriarty, USFWS)

Climate Change

Climate change, and specifically the anthropogenic warming trend presently underway, is recognized as a major extinction threat, particularly when combined with other threats such as habitat loss. Anthropogenic warming of the planet has been observed and is hypothesized to continue due to past and continuing emission of greenhouse gases, primarily carbon dioxide and methane, into the atmosphere caused by the burning of fossil fuels and deforestation. These gases decrease the degree to which Earth is able to radiate heat energy created by the sunlight that enters the atmosphere. The changes in climate and energy balance caused by increasing greenhouse gases are complex and our understanding of them depends on predictions generated from detailed computer models. Scientists generally agree the present warming trend is caused by humans and some of the likely effects include dramatic and dangerous climate changes in the coming decades. However, there is still debate and a lack of understanding about specific outcomes. Scientists disagree about the likely magnitude of the effects on extinction rates, with estimates ranging from 15 to 40 percent of species committed to extinction by 2050. Scientists do agree that climate change will alter regional climates, including rainfall and snowfall patterns, making habitats less hospitable to the species living in them. The warming trend will shift colder climates toward the north and south poles, forcing species to move with their adapted climate norms, but also to face habitat gaps along the way. The shifting ranges will impose new competitive regimes on species as they find themselves in contact with other species not present in their historic range. One such unexpected species contact is between polar bears and grizzly bears. Previously, these two species had separate ranges. Now, their ranges are overlapping and there are documented cases of these two species mating and producing viable offspring. Changing climates also

throw off the delicate timing adaptations that species have to seasonal food resources and breeding times. Scientists have already documented many contemporary mismatches to shifts in resource availability and timing.

Range shifts are already being observed: for example, on average, European bird species ranges have moved 91 km (56.5 mi) northward. The same study suggested that the optimal shift based on warming trends was double that distance, suggesting that the populations are not moving quickly enough. Range shifts have also been observed in plants, butterflies, other insects, freshwater fishes, reptiles, amphibians, and mammals.

Climate gradients will also move up mountains, eventually crowding species higher in altitude and eliminating the habitat for those species adapted to the highest elevations. Some climates will completely disappear. The rate of warming appears to be accelerated in the arctic, which is recognized as a serious threat to polar bear populations that require sea ice to hunt seals during the winter months: seals are the only source of protein available to polar bears. A trend to decreasing sea ice coverage has occurred since observations began in the mid-twentieth century. The rate of decline observed in recent years is far greater than previously predicted by climate models ([link]).

The effect of global warming can be seen in the continuing retreat of Grinnell Glacier. The mean annual temperature in Glacier National Park has increased 1.33°C since 1900. The

loss of a glacier results in the loss of summer meltwaters, sharply reducing seasonal water supplies and severely affecting local ecosystems. (credit: USGS, GNP Archives)

Finally, global warming will raise ocean levels due to meltwater from glaciers and the greater volume occupied by warmer water. Shorelines will be inundated, reducing island size, which will have an effect on some species, and a number of islands will disappear entirely. Additionally, the gradual melting and subsequent refreezing of the poles, glaciers, and higher elevation mountains—a cycle that has provided freshwater to environments for centuries—will be altered. This could result in an overabundance of salt water and a shortage of fresh water.

Section Summary

The core threats to biodiversity are human population growth and unsustainable resource use. To date, the most significant causes of extinction are habitat loss, introduction of exotic species, and overharvesting. Climate change is predicted to be a significant cause of extinction in the coming century. Habitat loss occurs through deforestation, damming of rivers, and other activities. Overharvesting is a threat particularly to aquatic species, but the taking of bush meat in the humid tropics threatens many species in Asia, Africa, and the Americas. Exotic species have been the cause of a number of extinctions and are especially damaging to islands and lakes. Exotic species' introductions are increasing because of the increased mobility of human populations and growing global trade and transportation. Climate change is forcing range changes that may lead to extinction. It is also affecting adaptations to the timing of resource availability that negatively affects species in seasonal environments. The impacts of climate change are currently greatest in the arctic. Global warming will also raise sea levels, eliminating some islands and reducing the area of all others.

Multiple Choice

Exercise:

Problem:

Converting a prairie to a farm field is an example of ________.

a. overharvesting
b. habitat loss
c. exotic species
d. climate change

Solution:

B

Exercise:

Problem:

Which two extinction risks may be a direct result of the pet trade?

a. climate change and exotic species introduction
b. habitat loss and overharvesting
c. overharvesting and exotic species introduction
d. habitat loss and climate change

Solution:

C

Exercise:

Problem:

What kind of ecosystem are exotic species especially threatening to?

a. deserts
b. marine ecosystems
c. islands

d. tropical forests

Solution:

C

Free Response

Exercise:

Problem:

Describe the mechanisms by which human population growth and resource use causes increased extinction rates.

Solution:

Human population growth leads to unsustainable resource use, which causes habitat destruction to build new human settlements, create agricultural fields, and so on. Larger human populations have also led to unsustainable fishing and hunting of wild animal populations. Excessive use of fossil fuels also leads to global warming.

Exercise:

Problem:

Explain what extinction threats a frog living on a mountainside in Costa Rica might face.

Solution:

The frog is at risk from global warming shifting its preferred habitat up the mountain. In addition, it will be at risk from exotic species, either as a new predator or through the impact of transmitted diseases such as chytridiomycosis. It is also possible that habitat destruction will threaten the species.

Glossary

bush meat
: a wild-caught animal used as food (typically mammals, birds, and reptiles); usually referring to hunting in the tropics of sub-Saharan Africa, Asia, and the Americas

chytridiomycosis
: a disease of amphibians caused by the fungus *Batrachochytrium dendrobatidis;* thought to be a major cause of the global amphibian decline

exotic species
: (also, invasive species) a species that has been introduced to an ecosystem in which it did not evolve

tragedy of the commons
: an economic principle that resources held in common will inevitably be over-exploited

white-nose syndrome
: a disease of cave-hibernating bats in the eastern United States and Canada associated with the fungus *Geomyces destructans*

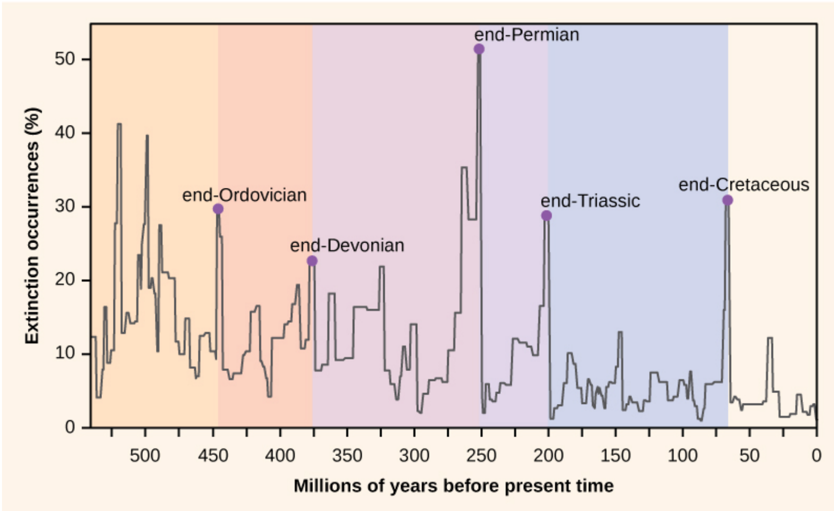
Preserving Biodiversity
By the end of this section, you will be able to:

- Describe biodiversity as the equilibrium of naturally fluctuating rates of extinction and speciation
- Explain the legislative framework for conservation
- Identify the factors important in conservation preserve design
- Identify examples of the effects of habitat restoration
- Identify the role of zoos in biodiversity conservation

Preserving biodiversity is an extraordinary challenge that must be met by greater understanding of biodiversity itself, changes in human behavior and beliefs, and various preservation strategies.

Change in Biodiversity through Time

The number of species on the planet, or in any geographical area, is the result of an equilibrium of two evolutionary processes that are ongoing: speciation and extinction. Both are natural "birth" and "death" processes of macroevolution. When speciation rates begin to outstrip extinction rates, the number of species will increase; likewise, the reverse is true when extinction rates begin to overtake speciation rates. Throughout the history of life on Earth, as reflected in the fossil record, these two processes have fluctuated to a greater or lesser extent, sometimes leading to dramatic changes in the number of species on the planet as reflected in the fossil record ([link]).



Extinction intensity as reflected in the fossil record has fluctuated throughout Earth's history. Sudden and dramatic losses of biodiversity, called mass extinctions, have occurred five times.

Paleontologists have identified five strata in the fossil record that appear to show sudden and dramatic (greater than half of all extant species disappearing from the fossil record) losses in biodiversity. These are called mass extinctions. There are many lesser, yet still dramatic, extinction events, but the five mass extinctions have attracted the most research into their causes. An argument can be made that the five mass extinctions are only the five most extreme events in a continuous series of large extinction events throughout the fossil record (since 542 million years ago). In most cases, the hypothesized causes are still controversial; in one, the most recent, the cause seems clear. The most recent extinction in geological time, about 65 million years ago, saw the disappearance of the dinosaurs and many other species. Most scientists now agree the cause of this extinction was the impact of a large asteroid in the present-day Yucatán Peninsula and the subsequent energy release and global climate changes caused by dust ejected into the atmosphere.

Recent and Current Extinction Rates

A sixth, or Holocene, mass extinction has mostly to do with the activities of *Homo sapiens*. There are numerous recent extinctions of individual species that are recorded in human writings. Most of these are coincident with the expansion of the European colonies since the 1500s.

One of the earlier and popularly known examples is the dodo bird. The dodo bird lived in the forests of Mauritius, an island in the Indian Ocean. The dodo bird became extinct around 1662. It was hunted for its meat by sailors and was easy prey because the dodo, which did not evolve with humans, would approach people without fear. Introduced pigs, rats, and dogs brought to the island by European ships also killed dodo young and eggs ([link]).

The dodo bird was hunted to extinction around 1662. (credit: Ed Uthman, taken in Natural History Museum, London, England)

Steller’s sea cow became extinct in 1768; it was related to the manatee and probably once lived along the northwest coast of North America. Steller’s sea cow was discovered by Europeans in 1741, and it was hunted for meat and oil. A total of 27 years elapsed between the sea cow’s first contact with Europeans and extinction of the species. The last Steller’s sea cow was killed in 1768. In another example, the last living passenger pigeon died in a zoo in Cincinnati, Ohio, in 1914. This species had once migrated in the millions but declined in numbers because of overhunting and loss of habitat through the clearing of forests for farmland.

These are only a few of the recorded extinctions in the past 500 years. The International Union for Conservation of Nature (IUCN) keeps a list of extinct and endangered species called the Red List. The list is not complete, but it describes 380 vertebrates that became extinct after 1500 AD, 86 of which were driven extinct by overhunting or overfishing.

Estimates of Present-day Extinction Rates

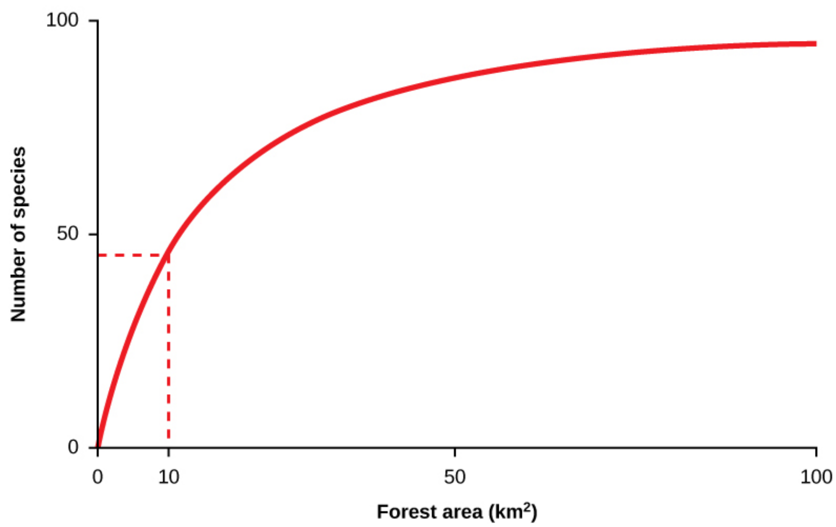
Estimates of **extinction rates** are hampered by the fact that most extinctions are probably happening without being observed. The extinction of a bird or mammal is often noticed by humans, especially if it has been hunted or used in some other way. But there are many organisms that are less noticeable to humans (not necessarily of less value) and many that are undescribed.

The background extinction rate is estimated to be about 1 per million species years (E/MSY). One “species year” is one species in existence for one year. One million species years could be one species persisting for one million years, or a million species persisting for one year. If it is the latter, then one extinction per million species years would be one of those million species becoming extinct in that year. For example, if there are 10 million species in existence, then we would expect 10 of those species to become extinct in a year. This is the background rate.

One contemporary extinction-rate estimate uses the extinctions in the written record since the year 1500. For birds alone, this method yields an estimate of 26 E/MSY, almost three times the background rate. However,

this value may be underestimated for three reasons. First, many existing species would not have been described until much later in the time period and so their loss would have gone unnoticed. Second, we know the number is higher than the written record suggests because now extinct species are being described from skeletal remains that were never mentioned in written history. And third, some species are probably already extinct even though conservationists are reluctant to name them as such. Taking these factors into account raises the estimated extinction rate to nearer 100 E/MSY. The predicted rate by the end of the century is 1500 E/MSY.

A second approach to estimating present-time extinction rates is to correlate species loss with habitat loss, and it is based on measuring forest-area loss and understanding species–area relationships. The **species-area relationship** is the rate at which new species are seen when the area surveyed is increased ([link]). Likewise, if the habitat area is reduced, the number of species seen will also decline. This kind of relationship is also seen in the relationship between an island’s area and the number of species present on the island: as one increases, so does the other, though not in a straight line. Estimates of extinction rates based on habitat loss and species–area relationships have suggested that with about 90 percent of habitat loss an expected 50 percent of species would become extinct. [link] shows that reducing forest area from 100 km^2 to 10 km^2, a decline of 90 percent, reduces the number of species by about 50 percent. Species–area estimates have led to estimates of present-day species extinction rates of about 1000 E/MSY and higher. In general, actual observations do not show this amount of loss and one explanation put forward is that there is a delay in extinction. According to this explanation, it takes some time for species to fully suffer the effects of habitat loss and they linger on for some time after their habitat is destroyed, but eventually they will become extinct. Recent work has also called into question the applicability of the species-area relationship when estimating the loss of species. This work argues that the species–area relationship leads to an overestimate of extinction rates. Using an alternate method would bring estimates down to around 500 E/MSY in the coming century. Note that this value is still 500 times the background rate.



A typical species-area curve shows the cumulative number of species found as more and more area is sampled. The curve has also been interpreted to show the effect on species numbers of destroying habitat; a reduction in habitat of 90 percent from 100 km^2 to 10 km^2 reduces the number of species supported by about 50 percent.

Note:
Concept in Action

Go to this website for an interactive exploration of endangered and extinct species, their ecosystems, and the causes of their endangerment or extinction.

Conservation of Biodiversity

The threats to biodiversity at the genetic, species, and ecosystem levels have been recognized for some time. In the United States, the first national park with land set aside to remain in a wilderness state was Yellowstone Park in 1890. However, attempts to preserve nature for various reasons have occurred for centuries. Today, the main efforts to preserve biodiversity involve legislative approaches to regulate human and corporate behavior, setting aside protected areas, and habitat restoration.

Changing Human Behavior

Legislation has been enacted to protect species throughout the world. The legislation includes international treaties as well as national and state laws. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) treaty came into force in 1975. The treaty, and the national legislation that supports it, provides a legal framework for preventing "listed" species from being transported across nations' borders, thus protecting them from being caught or killed in the first place when the purpose involves international trade. The listed species that are protected to one degree or another by the treaty number some 33,000. The treaty is limited in its reach because it only deals with international movement of organisms or their parts. It is also limited by various countries' ability or willingness to enforce the treaty and supporting legislation. The illegal trade in organisms and their parts is probably a market in the hundreds of millions of dollars.

Within many countries there are laws that protect endangered species and that regulate hunting and fishing. In the United States, the Endangered Species Act was enacted in 1973. When an at-risk species is listed by the Act, the U.S. Fish & Wildlife Service is required by law to develop a management plan to protect the species and bring it back to sustainable numbers. The Act, and others like it in other countries, is a useful tool, but it suffers because it is often difficult to get a species listed, or to get an effective management plan in place once a species is listed. Additionally, species may be controversially taken off the list without necessarily having

had a change in their situation. More fundamentally, the approach to protecting individual species rather than entire ecosystems (although the management plans commonly involve protection of the individual species' habitat) is both inefficient and focuses efforts on a few highly visible and often charismatic species, perhaps at the expense of other species that go unprotected.

The Migratory Bird Treaty Act (MBTA) is an agreement between the United States and Canada that was signed into law in 1918 in response to declines in North American bird species caused by hunting. The Act now lists over 800 protected species. It makes it illegal to disturb or kill the protected species or distribute their parts (much of the hunting of birds in the past was for their feathers). Examples of protected species include northern cardinals, the red-tailed hawk, and the American black vulture.

Global warming is expected to be a major driver of biodiversity loss. Many governments are concerned about the effects of anthropogenic global warming, primarily on their economies and food resources. Since greenhouse gas emissions do not respect national boundaries, the effort to curb them is an international one. The international response to global warming has been mixed. The Kyoto Protocol, an international agreement that came out of the United Nations Framework Convention on Climate Change that committed countries to reducing greenhouse gas emissions by 2012, was ratified by some countries, but spurned by others. Two countries that were especially important in terms of their potential impact that did not ratify the Kyoto protocol were the United States and China. Some goals for reduction in greenhouse gasses were met and exceeded by individual countries, but, worldwide, the effort to limit greenhouse gas production is not succeeding. The intended replacement for the Kyoto Protocol has not materialized because governments cannot agree on timelines and benchmarks. Meanwhile, the resulting costs to human societies and biodiversity predicted by a majority of climate scientists will be high.

As already mentioned, the non-profit, non-governmental sector plays a large role in conservation effort both in North America and around the world. The approaches range from species-specific organizations to the broadly focused IUCN and Trade Records Analysis of Flora and Fauna in

Commerce (TRAFFIC). The Nature Conservancy takes a novel approach. It purchases land and protects it in an attempt to set up preserves for ecosystems. Ultimately, human behavior will change when human values change. At present, the growing urbanization of the human population is a force that mitigates against valuing biodiversity, because many people no longer come in contact with natural environments and the species that inhabit them.

Conservation in Preserves

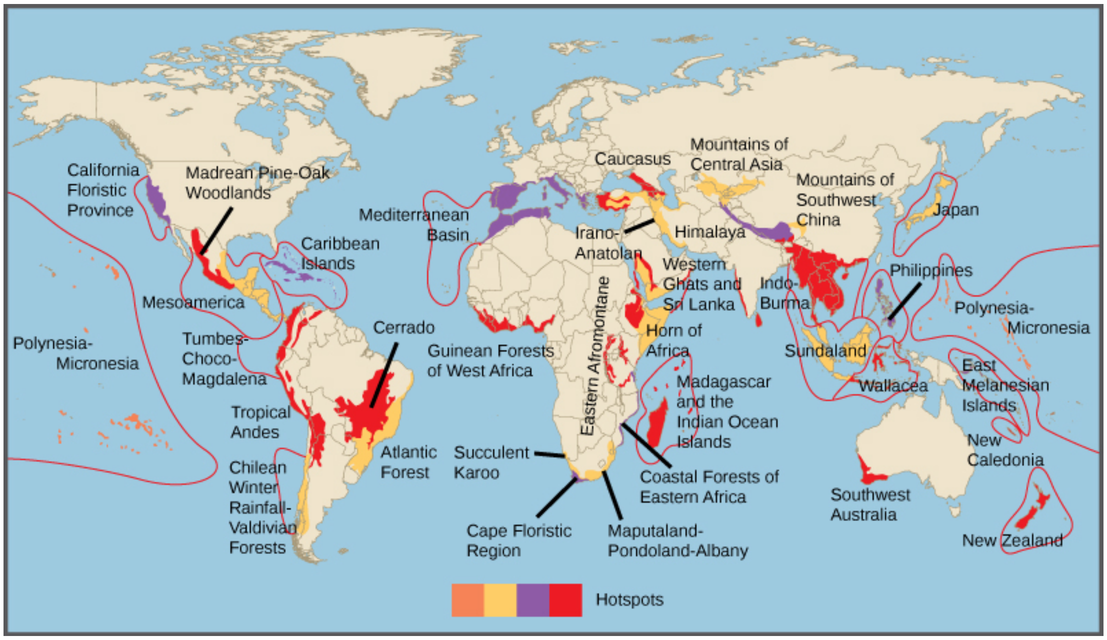
Establishment of wildlife and ecosystem preserves is one of the key tools in conservation efforts ([link]). A preserve is an area of land set aside with varying degrees of protection for the organisms that exist within the boundaries of the preserve. Preserves can be effective for protecting both species and ecosystems, but they have some serious drawbacks.

National parks, such as Grand Teton National Park in Wyoming, help conserve biodiversity. (credit: Don DeBold)

A simple measure of success in setting aside preserves for biodiversity protection is to set a target percentage of land or marine habitat to protect.

However, a more detailed preserve design and choice of location is usually necessary because of the way protected lands are allocated and how biodiversity is distributed: protected lands tend to contain less economically valuable resources rather than being set aside specifically for the species or ecosystems at risk. In 2003, the IUCN World Parks Congress estimated that 11.5 percent of Earth’s land surface was covered by preserves of various kinds. This area is greater than previous goals; however, it only represents 9 out of 14 recognized major biomes and research has shown that 12 percent of all species live outside preserves; these percentages are much higher when threatened species are considered and when only high quality preserves are considered. For example, high quality preserves include only about 50 percent of threatened amphibian species. The conclusion must be that either the percentage of area protected must be increased, the percentage of high quality preserves must be increased, or preserves must be targeted with greater attention to biodiversity protection. Researchers argue that more attention to the latter solution is required.

A **biodiversity hotspot** is a conservation concept developed by Norman Myers in 1988. Hotspots are geographical areas that contain high numbers of endemic species. The purpose of the concept was to identify important locations on the planet for conservation efforts, a kind of conservation triage. By protecting hotspots, governments are able to protect a larger number of species. The original criteria for a hotspot included the presence of 1500 or more species of endemic plants and 70 percent of the area disturbed by human activity. There are now 34 biodiversity hotspots ([link]) that contain large numbers of endemic species, which include half of Earth’s endemic plants.



Conservation International has identified 34 biodiversity hotspots. Although these cover only 2.3 percent of the Earth's surface, 42 percent of the terrestrial vertebrate species and 50 percent of the world's plants are endemic to those hotspots.

There has been extensive research into optimal preserve designs for maintaining biodiversity. The fundamental principles behind much of the research have come from the seminal theoretical work of Robert H. MacArthur and Edward O. Wilson published in 1967 on island biogeography.[footnote] This work sought to understand the factors affecting biodiversity on islands. Conservation preserves can be seen as "islands" of habitat within "an ocean" of non-habitat. In general, large preserves are better because they support more species, including species with large home ranges; they have more core area of optimal habitat for individual species; they have more niches to support more species; and they attract more species because they can be found and reached more easily.

Robert H. MacArthur and Edward O. Wilson, E. O., *The Theory of Island Biogeography* (Princeton, N.J.: Princeton University Press, 1967).

Preserves perform better when there are partially protected buffer zones around them of suboptimal habitat. The buffer allows organisms to exit the boundaries of the preserve without immediate negative consequences from hunting or lack of resources. One large preserve is better than the same area of several smaller preserves because there is more core habitat unaffected by less hospitable ecosystems outside the preserve boundary. For this same reason, preserves in the shape of a square or circle will be better than a preserve with many thin “arms.” If preserves must be smaller, then providing wildlife corridors between them so that species and their genes can move between the preserves; for example, preserves along rivers and streams will make the smaller preserves behave more like a large one. All of these factors are taken into consideration when planning the nature of a preserve before the land is set aside.

In addition to the physical specifications of a preserve, there are a variety of regulations related to the use of a preserve. These can include anything from timber extraction, mineral extraction, regulated hunting, human habitation, and nondestructive human recreation. Many of the decisions to include these other uses are made based on political pressures rather than conservation considerations. On the other hand, in some cases, wildlife protection policies have been so strict that subsistence-living indigenous populations have been forced from ancestral lands that fell within a preserve. In other cases, even if a preserve is designed to protect wildlife, if the protections are not or cannot be enforced, the preserve status will have little meaning in the face of illegal poaching and timber extraction. This is a widespread problem with preserves in the tropics.

Some of the limitations on preserves as conservation tools are evident from the discussion of preserve design. Political and economic pressures typically make preserves smaller, never larger, so setting aside areas that are large enough is difficult. Enforcement of protections is also a significant issue in countries without the resources or political will to prevent poaching and illegal resource extraction.

Climate change will create inevitable problems with the location of preserves as the species within them migrate to higher latitudes as the habitat of the preserve becomes less favorable. Planning for the effects of

global warming on future preserves, or adding new preserves to accommodate the changes expected from global warming is in progress, but will only be as effective as the accuracy of the predictions of the effects of global warming on future habitats.

Finally, an argument can be made that conservation preserves reinforce the cultural perception that humans are separate from nature, can exist outside of it, and can only operate in ways that do damage to biodiversity. Creating preserves reduces the pressure on human activities outside the preserves to be sustainable and non-damaging to biodiversity. Ultimately, the political, economic, and human demographic pressures will degrade and reduce the size of conservation preserves if the activities outside them are not altered to be less damaging to biodiversity.

Note:
Concept in Action

Check out this interactive global data system of protected areas. Review data about specific protected areas by location or study statistics on protected areas by country or region.

Habitat Restoration

Habitat restoration holds considerable promise as a mechanism for maintaining or restoring biodiversity. Of course once a species has become extinct, its restoration is impossible. However, restoration can improve the biodiversity of degraded ecosystems. Reintroducing wolves, a top predator, to Yellowstone National Park in 1995 led to dramatic changes in the

ecosystem that increased biodiversity. The wolves ([link]) function to suppress elk and coyote populations and provide more abundant resources to the guild of carrion eaters. Reducing elk populations has allowed revegetation of riparian (the areas along the banks of a stream or river) areas, which has increased the diversity of species in that habitat. Suppression of coyotes has increased the species previously suppressed by this predator. The number of species of carrion eaters has increased because of the predatory activities of the wolves. In this habitat, the wolf is a keystone species, meaning a species that is instrumental in maintaining diversity within an ecosystem. Removing a keystone species from an ecological community causes a collapse in diversity. The results from the Yellowstone experiment suggest that restoring a keystone species effectively can have the effect of restoring biodiversity in the community. Ecologists have argued for the identification of keystone species where possible and for focusing protection efforts on these species. It makes sense to return the keystone species to the ecosystems where they have been removed.

This photograph shows the Gibbon wolf pack in Yellowstone National Park, March 1, 2007. Wolves have been identified as a keystone species. (credit: Doug Smith, NPS)

Other large-scale restoration experiments underway involve dam removal. In the United States, since the mid-1980s, many aging dams are being considered for removal rather than replacement because of shifting beliefs about the ecological value of free-flowing rivers. The measured benefits of dam removal include restoration of naturally fluctuating water levels (often the purpose of dams is to reduce variation in river flows), which leads to increased fish diversity and improved water quality. In the Pacific Northwest, dam removal projects are expected to increase populations of salmon, which is considered a keystone species because it transports nutrients to inland ecosystems during its annual spawning migrations. In other regions, such as the Atlantic coast, dam removal has allowed the return of other spawning anadromous fish species (species that are born in fresh water, live most of their lives in salt water, and return to fresh water to spawn). Some of the largest dam removal projects have yet to occur or have happened too recently for the consequences to be measured. The large-scale ecological experiments that these removal projects constitute will provide valuable data for other dam projects slated either for removal or construction.

The Role of Zoos and Captive Breeding

Zoos have sought to play a role in conservation efforts both through captive breeding programs and education ([link]). The transformation of the missions of zoos from collection and exhibition facilities to organizations that are dedicated to conservation is ongoing. In general, it has been recognized that, except in some specific targeted cases, captive breeding programs for endangered species are inefficient and often prone to failure when the species are reintroduced to the wild. Zoo facilities are far too limited to contemplate captive breeding programs for the numbers of species that are now at risk. Education, on the other hand, is a potential positive impact of zoos on conservation efforts, particularly given the global trend to urbanization and the consequent reduction in contacts between people and wildlife. A number of studies have been performed to

look at the effectiveness of zoos on people’s attitudes and actions regarding conservation; at present, the results tend to be mixed.

Zoos and captive breeding programs help preserve many endangered species, such as this golden lion tamarin. (credit: Garrett Ziegler)

Section Summary

Five mass extinctions with losses of more than 50 percent of extant species are observable in the fossil record. Recent extinctions are recorded in written history and are the basis for one method of estimating contemporary extinction rates. The other method uses measures of habitat loss and species-area relationships. Estimates of contemporary extinction rates vary

but are as high as 500 times the background rate, as determined from the fossil record, and are predicted to rise.

There is a legislative framework for biodiversity protection. International treaties such as CITES regulate the transportation of endangered species across international borders. Legislation within individual countries protecting species and agreements on global warming have had limited success; there is at present no international agreement on targets for greenhouse gas emissions. In the United States, the Endangered Species Act protects listed species but is hampered by procedural difficulties and a focus on individual species. The Migratory Bird Act is an agreement between Canada and the United States to protect migratory birds. The non-profit sector is also very active in conservation efforts in a variety of ways.

Conservation preserves are a major tool in biodiversity protection. Presently, 11 percent of Earth’s land surface is protected in some way. The science of island biogeography has informed the optimal design of preserves; however, preserves have limitations imposed by political and economic forces. In addition, climate change will limit the effectiveness of present preserves in the future. A downside of preserves is that they may lessen the pressure on human societies to function more sustainably outside the preserves.

Habitat restoration has the potential to restore ecosystems to previous biodiversity levels before species become extinct. Examples of restoration include reintroduction of keystone species and removal of dams on rivers. Zoos have attempted to take a more active role in conservation and can have a limited role in captive breeding programs. Zoos also have a useful role in education.

Multiple Choice

Exercise:

Problem:

Certain species of parrot cannot be brought to the United States to be sold as pets. What is the name of the legislation that makes this illegal?

a. Red List
b. Migratory Bird Act
c. CITES
d. Endangered Species Act (ESA)

Solution:

C

Exercise:

Problem:

What is the name of the first international agreement on climate change?

a. Red List
b. Montreal Protocol
c. International Union for the Conservation of Nature (IUCN)
d. Kyoto Protocol

Solution:

D

Free Response

Exercise:

Problem: Describe two considerations in conservation preserve design.

Solution:

Larger preserves will contain more species. Preserves should have a buffer around them to protect species from edge effects. Preserves that are round or square are better than preserves with many thin arms.

Exercise:

Problem:

Describe what happens to an ecosystem when a keystone species is removed.

Solution:

Many species will disappear from the ecosystem when a keystone species is removed.

Glossary

biodiversity hotspot
: a concept originated by Norman Myers to describe a geographical region with a large number of endemic species and a large percentage of degraded habitat

extinction rate
: the number of species becoming extinct over time, sometimes defined as extinctions per million species–years to make numbers manageable (E/MSY)

species-area relationship
: the relationship between area surveyed and number of species encountered; typically measured by incrementally increasing the area of a survey and determining the cumulative numbers of species

The Periodic Table of Elements